NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A STATISTICALLY BASED TRAINING DIAGNOSTIC TOOL FOR MARINE AVIATION

by

Francis M. Pascucci

June 2014

Thesis Co-Advisors: Samuel E. Buttrey
Joseph Sullivan

This thesis was performed at the MOVES Institute
Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2014	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE A STATISTICALLY BASED TRAINING DIAGNOSTIC TOOL FOR MARINE AVIATION			**5. FUNDING NUMBERS**
6. AUTHOR(S) Francis M. Pascucci			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB protocol number NPS.2013.0090-IR-EP7-A.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

This work focused on the design of a graphical user interface to improve instructional design models and decision support for Marine aviation training. Trainee performance data was collected, analyzed, and compared the results of a survey of instructor pilots to find correlations between the scores assigned and opinions on the critical items identified by instructors. This information was used to inform the design of a system that provides leadership with trainee trends in visual form. Such a system could allow for early training interventions for those who struggle and better training management for those who are excelling.

Although this thesis focused on the aviation domain, this methodology could be generalized to any U.S. Marine Corps or military training evaluation system using a criteria-referenced performance rating system. The sample data did not provide sufficient statistical evidence to predict future performance; however, it was sufficient to provide a meaningful visual representation of performance trends. The results gained in the analysis allowed for recommendations on changes to the current evaluation system and improvements to the technologies used to inform decision makers. A prototype of the designed graphical user interface is presented.

14. SUBJECT TERMS Training evaluation, decision support, instructional design			**15. NUMBER OF PAGES** 151
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

A STATISTICALLY BASED TRAINING DIAGNOSTIC TOOL FOR MARINE AVIATION

Francis M. Pascucci
Captain, United States Marine Corps
B.S., United States Naval Academy, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MODELING VIRTUAL ENVIRONMENTS AND SIMULATION

from the

NAVAL POSTGRADUATE SCHOOL
June 2014

Author: Francis M. Pascucci

Approved by: Samuel E. Buttrey
Thesis Co-Advisor

Joseph Sullivan
Thesis Co-Advisor

Christian Darken
Chair, Modeling Virtual Environments and Simulation

Peter Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This work focused on the design of a graphical user interface to improve instructional design models and decision support for Marine aviation training. Trainee performance data was collected, analyzed, and compared the results of a survey of instructor pilots to find correlations between the scores assigned and opinions on the critical items identified by instructors. This information was used to inform the design of a system that provides leadership with trainee trends in visual form. Such a system could allow for early training interventions for those who struggle and better training management for those who are excelling.

Although this thesis focused on the aviation domain, this methodology could be generalized to any U.S. Marine Corps or military training evaluation system using a criteria-referenced performance rating system. The sample data did not provide sufficient statistical evidence to predict future performance; however, it was sufficient to provide a meaningful visual representation of performance trends. The results gained in the analysis allowed for recommendations on changes to the current evaluation system and improvements to the technologies used to inform decision makers. A prototype of the designed graphical user interface is presented.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACE	aviation combat element
ACPM	aviation career progression model
ADDIE	analyze design develop implement evaluate
AH	attack helicopter
AI	air interdiction
API	aviation preflight indoctrination
APR	aviation performance record
ASPT	assault support
ATD	aviation training division
ATF	aviation tracking file
ATJ	aviation training jacket
ATS	aviation training system
BARS	behaviorally anchored rating scale
BIP	basic instructor pilot
CaRBS	classification and ranking belief simplex
CAS	close air support
CAL	confined area landing
CRM	crew resource management
CRP	current readiness program
DACMI	defensive air combat maneuver instructor
DOSS	department of standardization and safety
DRRS	Defense Readiness Reporting System
ESC	escort
FAC(A)I	forward air controller (airborne) instructor
FAM	familiarization
FITREP	fitness report
FLSE	flight leadership standardization evaluator
FORM	formation
FRS	fleet replacement squadron
FRSI	fleet replacement squadron instructor

IFS	initial flight screening
IP	instructor pilot
IUT	instructor under training
MAGTF	Marine air ground task force
MATSS	Marine aviation training support site
MET	mission essential task
METL	mission essential task list
MDG	maneuver description guide
MOS	military occupational specialty
MSHARP	Marine Sierra Hotel Aviation Reporting Program
NAVMC	Navy and Marine Corps
NSI	night systems instructor
NSFI	night systems familiarization instructor
NSS	Navy Standard Score
NVD	night vision device
PROMETHEE	preference ranking organization method for enrichment evaluation
PUI	pilot under instruction
RAC	replacement aircrew
RAT	requires additional training
SNA	student naval aviator
SNFO	student naval flight officer
SWD	specific weapons delivery
TERFI	terrain flight instructor
T/M/S	type/model/series
T&R	training and readiness
TSI	tactical simulator instructor
UH	utility helicopter
QCA	qualitative comparative analysis
UX	user experience
WTO	weapons training officer

I. INTRODUCTION

Within the community of naval aviation, pilots and naval flight officers undergo a thorough and extensive training program before arriving at their first operational squadron. Despite having spent approximately 18 to 24 months being trained to achieve the designation as a naval aviator or flight officer, their training continues throughout their time in the operational environment. This training is focused on teaching designated aviators how to tactically employ their aircraft across the full spectrum of operations.

The instructors conducting each training syllabus event are required to complete an aviation tracking file (ATF) that records the pilot under instruction's (PUI) performance via an enumerated list of metrics determined by the *Training and Readiness* (T&R) *Manual*. The T&R manual mandates that ATFs be completed for any initial event completed by aviators during their initial accession of skills, during a refresher syllabus, or while executing a series conversion (Headquarters United States Marine Corps, 2011a, p. 2-10). The T&R manual is silent on exactly how instructors should fill the ATF out, in terms of selecting grades and writing comments. The ATF provides feedback to the trainee and performance information to other instructors and the unit's leadership on how that individual pilot is performing and progressing through the designated syllabus. This information is reviewed by several levels of stakeholders within the command. These stakeholders include the Squadron Department of Standardization and Safety (DOSS), to ensure events are conducted safely; the operations officer, to ensure that events are completed for pilot progression and to maintain and build unit-level personnel proficiency requirements; the executive officer; and the commanding officer, as well as instructors, who to some degree, rely on the information to profile aviators in a training syllabus.

A considerable amount of time and effort is put into writing ATFs, discussing which pilots in a training syllabus are succeeding, and which are not, and determining what training items should be stressed, due to deficiencies or weak points among the entire cadre of aviators in the squadron. Despite a great deal of information available and

accessible through ATFs written by instructors, aviation units have mostly relied on informal discussions, which provided anecdotal evidence to make these decisions of how to better train individuals and the squadron.

The under-utilization of ATF data as a resource to better inform decision-making is a result of a combination of factors. First and foremost, ATFs are contained within each individual aviator's aircrew performance record (APR), which consists of a five-part file folder containing paper copies of each ATF written for that particular individual. These ATFs are not tracked outside of the squadron in any form and official records exist solely in the paper format within the APR. Additionally, due to time constraints placed on instructors within the unit, the full APR is rarely taken into account by trainers. Instead, the most recent ATFs might be scanned for strengths and weaknesses of the trainee, and the instructors with whom the trainee flew with might be consulted to discuss the individual's performance. Furthermore, in discussions held among senior leadership and the instructor cadre, opinions are solicited on the progression and performance of each individual trainee. In general, if the individual has completed his or her most recent flights with no glaring deficiencies, he or she is generally accepted as performing satisfactorily. These instructor meetings are usually attended by all available instructors, but often not the full instructor cadre due to other commitments (e.g., scheduled flights, medical appointments). This results in some discussion of trainees' recent performance not being addressed if the instructor who most recently flew with that individual is absent or fails to communicate relevant issues to the group that arose during a flight.

A. SYSTEM PURPOSE

Marine aviation currently relies on manual review of ATF data and discussions held amongst instructors to determine the level of trainee performance. Statistical methods can be applied to the existing data to help quantify trainee performance. Using these methods a better understanding can be gained by stakeholders on the performance of individual trainees and the instructional system. Furthermore, the development of a tool that increases the robustness of the instructional system has the potential to improve readiness and reduce costs. The primary purpose of the Statistically Based Training

Diagnostic Tool for Marine aviation is to aid the stakeholders in assessing the performance of aviators within the operational environment. The stakeholders include trainees, instructor cadre, the squadron leadership and potentially leadership at the group level and above. By having a tool that enables these stakeholders to visualize and understand trends of individuals and groups of trainees, training can be tailored to address deficiencies and highlight proficiencies. The existence of this tool will provide an option for instructors and leadership to understand the wealth of information regarding pilot training that hours of time are spent creating. When this information is readily available the potential for a more effectively and efficiently trained force exists. The potential also exists to enhance senior leadership knowledge of how well subordinate units are trained, in contrast to only knowing the qualification level to which they are trained.

B. PROBLEM STATEMENT

The current utilization of the recorded training documentation does not include any empirical analysis regarding the numerical scores or of the subjective comments that are provided by instructors following each training event, including both those completed in the simulator and in the actual aircraft. No data has been collected on identifying critical performance items that identify difficulties being experienced by PUIs, nor has a method been developed to address the summarization and utilization of this data. Presently no methods exist to efficiently observe and understand the relevance of empirical performance information of individual aviators within Marine aviation. Decision makers need convenient access to performance data so that unit leadership can better understand the level at which personnel are being trained.

C. RELEVANCE TO THE DEPARTMENT OF DEFENSE

Currently training performance data is not objectively and empirically analyzed within operational Marine aviation units preparing warfighters to execute their war-time duties. This thesis will explore the capability to provide trainers and leadership with a data-driven training diagnostic tool to facilitate greater effectiveness and efficiency for individual warfighters and for the collective unit. In addition, recognizing subtle developmental training deficiencies can provide increased safety and reduced costs due to

loss and damage. Marine Corps Training and Education Science and Technology Objective-2: Small Unit Learning and Performance Assessment in the *USMC Science and Technology Strategic Plan 2012* calls for "valid scientific products and affordable technologies to unobtrusively assess and predict performance" (Office of the Deputy Commandant for Combat Development Integration, 2012, p. 34). Future application for this work could be seen within all types of units, to measure and adjust training programs to better meet the needs of trainees.

D. RESEARCH QUESTIONS

This thesis will be guided by the following questions:

- Can an analysis tool be created that provides an interface to display training information providing actionable metrics that allow for training program intervention and remediation using existing performance models to identify strengths, weaknesses and trends among trainees?
- Do numerical grades and/or comments on specific graded items predict future performance success or failure?
- If correlations exist, can they be identified mid-syllabus, when the training syllabus can be adjusted or supplemented to remedy deficiencies?

E. SCOPE AND LIMITATIONS

This thesis involves the collection of training data from operational squadrons, analysis of that data, and the collection of survey data that exposes criteria that operational instructors deem most critical in evaluating a PUI's progression and development within their professional domain. The collection of this data is driving the development of a prototype of a system that can provide a summarization of PUI performance that highlights critical performance measures and is presented in an intuitive and understandable manner. This prototype will not be a fully operational system, but rather a recommendation for a fully implementable design.

F. THESIS ORGANIZATION

This thesis is organized into five chapters. Chapter I introduces the motivation for this research effort. It outlines the purpose for pursuing further understanding of evaluation of aviation trainees, which can be generalized across other military domains.

The interest in the efforts of the Department of Defense is addressed. Specific research questions that this thesis attempts to answer are stated and, finally, the scope and limitations of this research effort are discussed.

Chapter II provides a background of the research domain and an in-depth review of key concepts and theories that pertain to this effort. It contains information regarding the naval aviation training progression, the Marine Corps training methodology, instructional design, evaluation methodologies, and decision support systems and their design.

Chapter III describes the methodology adopted to conduct the research and attempt to answer the given research questions in the given domain.

Chapter IV consists of the analysis of the two data sets collected for this research and the application of these results to model a decision support tool. The first data set is comprised of aviation training form data containing graded items intended to provide pilot performance information. The second set of data consists of survey results obtained on instructor pilot opinions of the aviation training form and current method in use to evaluate trainees.

Chapter V contains the conclusions and recommendations from this research effort, as well as discussion of future research efforts that could be conducted in this domain.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

Since 1912, when the first Marine officer reported to Annapolis, Maryland for initial flight training, the United States Marine Corps has been linked to naval aviation (Mersky, 1983). Today, Marine Corps aviators train side by side with their Navy counterparts in the initial accession in the aviation pipeline. The initial training undergone as a student naval aviator permeates all of an aviator's future training when preparing for combat missions in support of operations conducted by the United States Department of Defense. As such, the training is intended to be thorough and extensive to produce capable combat aviators. The naval aviation training pipeline has undergone a number of changes and transitions in adopting new technologies and methodologies over the years to continue producing high-quality aviators. Today, the training pipeline is a complex system that ultimately results in designated aviators continuing their training and development throughout their career.

Both the military and civilian aviation domains have conducted research to investigate predictive markers for naval aviator performance. It is a primary concern in the military domain based on improved safety as well as considerable monetary savings. Shannon and Waag (1972) attempted to isolate the critical skill sets and procedures within the West Coast Replacement Air Group, now known as the fleet replacement squadron (FRS), to determine predictive measures of both intermediate stage grading and final grading. This study found that the selected measures were highly correlated with the results from a similar study completed utilizing the East Coast FRS and the same critical items (Shannon & Waag, 1972). Rickus and Berkshire (1968) attempted to address the criterion for prediction of aviators combat performance, making a distinction between the early stages of flight training and mission oriented activities. Another study identified 10 specific behaviors that could be utilized as predictive of aviator success in early flight training (Stanley Jr., 1973). Hunter and Burke (2009) conducted a meta-analysis of published research pertaining to predicting pilot performance and addressing the validity of the several criterion identified as predictive. More recently there have also been efforts to utilize neural networks and multiple regression to predict pilot success (Griffin, 1998).

This research can be extended by looking at current Naval Aviator performance and subject matter expert opinion regarding the critical factors that comprise the grades being received by trainees. The use of predictive measures in this thesis serves to provide a means to pin-point the shortcomings to allow for training interventions and prevent future failures or increase levels of success.

A. NAVAL AVIATION TRAINING PROGRESSION

The naval aviation training pipeline consists of undergraduate and graduate level aviation training, culminating in the designation of a prospective aviator as either a naval aviator or naval flight officer. Throughout the training program prospective aviators are continuously evaluated using a number of different methods depending on the phase. Prospective naval aviators begin their training in the Initial Flight Screening (IFS) program. This program, consisting of 25 flight hours in civilian fixed-wing aircraft, was implemented to expose selected prospective student naval aviators (SNA) and student naval flight officers (SNFO) with no prior aviation experience to the aeronautical environment, and to identify students who no longer desire to pursue a career in military aviation after this exposure. Completion of the IFS program is a requirement for SNAs and SNFOs prior to entering the Aviation Preflight Indoctrination (API) phase of Naval Aviation Training. Having completed with IFS, commissioned naval officers proceed to Pensacola, Florida to enter API. API consists of a six-week period of instruction covering the basics of engineering, aerodynamics, weather, navigation, flight rules and regulations, aviation physiology, and water survival.

Following the successful completion of API, SNAs and SNFOs branch into their respective pipelines, which differ for pilots and flight officers. From this point forward we will focus on the training of SNAs. The next phase of training for SNAs is Primary Flight Training. The primary phase of training is conducted at NAS Whiting Field in Milton, Florida, NAS Corpus Christi, Texas, or Vance AFB in Enid, Oklahoma. The students at the Navy locations undergo an approximately 22-week course of instruction learning airmanship in either the T-34C Turbomentor or the T-6A Texan II turbo-prop, fixed-wing aircraft. During this training SNAs are evaluated using the Multi-Service Pilot

Training System (MPTS), which is a "two phased, pilot training curriculum utilizing Course Training Standards and Maneuver Item Files to identify acceptable levels of training performance" (H-3, Naval Air Training Command, 2007). It is important to note that at the completion of this phase SNAs have the opportunity to express their preferences as to what type of platform they wish to fly in the operational fleet. They may choose tactical jets, rotary-wing (helicopters), multi-engine platforms, or, for the Marine SNAs, tilt-rotor aircraft (MV-22 Osprey). Depending on the needs of their respective service, their performance in the primary phase, and their preferences, SNAs are assigned to either intermediate jet training, intermediate rotary-wing training (tilt-rotor selectees), advanced maritime training (multi-engine selectees), or advanced rotary wing training. Students who complete intermediate jet training continue to either advanced strike training or advanced E-2/C-2A training, and earn their designation as a naval aviator at the completion of this advanced training. Students who are assigned to advanced rotary-wing training or advanced maritime training earn the naval aviator designation at the completion of that phase.

Following designation as a naval aviator, naval officers report to the FRS. At the FRS aviators are trained in their respective operational aircraft (e.g., F/A-18, AH-1W, SH-60, etc.). While training syllabi for the individual platforms vary in the number of flight events required, the main focus for all FRS activities is to train aviators in the flight characteristics, emergency procedures, and operation of their respective platform. While some tactical flight exposure is conducted during FRS training, the majority of tactical flight training occurs in operational squadrons. It is in the operational squadron where aviators are initially trained in the tactical employment of their aircraft.

B. SYSTEMS APPROACH TO TRAINING

The systems approach to training is a method in which a systematic method is applied to develop the entirety of the training progression to ensure the end-state is achievable in an effective and efficient manner. The *Marine Corps Systems Approach to Training Manual* states, "The goal of Marine Corps instruction is to develop performance-based, criterion-referenced instruction that promotes student transfer of

learning from the instructional setting to the job" (U.S. Marine Corps, 2004, p. ii). Gagné and Briggs (1979) point out that the intent of instructional systems design "attempts to bring systematic knowledge of the learning process to bear on the design of instruction," (p. 20). The *Systems Approach to Training Manual* follows Gagné and Briggs's (1979) instructional design model while also making reference to Bloom (1956). The intent of the systems approach to training is to leverage each stage of instruction to harness human learning capabilities with delivery methods, to increase effectiveness and efficiency. The Marine Corps' adoption of the Aviation Training System (ATS) is an attempt to fully implement the systems approach to training in Marine aviation (Fenwick, 2010). According to the *Aviation Training and Readiness Program*, "The purpose of ATS is to develop and maintain a fully integrated training system across all of Marine Aviation," (Headquarters United States Marine Corps, 2011, p. 2-4). The ATS is supposed to leverage Marine aviation training support sites (MATSS) at each Marine air station primarily to increase efficiency with regards to asset (simulator) utilization and standardization of training.

C. THE TRAINING AND READINESS PROGRAM

The Navy and Marine Corps (NAVMC) Aviation Training and Readiness Program provides the foundation for the implementation and administration of training programs, and the methods by which to measure and monitor their effectiveness. For Marine aviation, *NAVMC 3500.14C* is the governing document that outlines the requirements for all aviation training activities in the Marine Corps. The *Aviation Training and Readiness Program Manual* states the following:

> The Marine Aviation Training and Readiness (T&R) Program provides the Marine Air-Ground Task Force (MAGTF) commander with an Aviation Combat Element (ACE) capable of executing the six functions of Marine Aviation. The T&R Program is the fundamental tool used by commanders to construct, attain, and maintain effective training programs and is the foundation for the Aviation Training System (ATS). (Headquarters United States Marine Corps, 2011b)

The *Aviation Training and Readiness Program Manual* requires that each operational platform have its own specific training and readiness program (Headquarters

United States Marine Corps, 2011b). This thesis will focus on the training policies and rules of conduct, the separate phases of training, and the management and evaluation of readiness.

1. The Core Competency Model

The core competency model, also referred to as the core model, is the standardized foundation on that all platform specific Training and Readiness programs are built upon (Headquarters United States Marine Corps, 2011, p. 2-3). The model is separated into phases that are related to the mission requirements of the particular platform community. The phases are delineated in Table 1.

Phase	TERM	DEFINITION
1000	Core Skill Introduction	Entry level training required to receive or be eligible for assignment of a primary MOS. Includes such training as systems / equipment, operations familiarization, initial crew procedures, and initial exposure to core skills.
2000	Core Skill	Fundamental, environmental, or conditional capabilities required to perform basic functions. These basic functions serve as tactical enablers that allow crews to progress to the more complex Mission Skills.
3000	Mission Skill	Mission Skills enable a unit to execute a specific MET. They are comprised of advanced event(s) that are focused on MET performance and draw upon the knowledge, abilities, and situational awareness developed during Core Skill training.
4000 4500	Core Plus Skill Mission Plus	Training events that can be theater specific or that have a low likelihood of occurrence. They may be fundamental, environmental, or conditional capabilities required to perform basic functions.
5000	Instructor Training	Instructor training events.
6000	Requirements, Certifications, Qualifications, and Designations (R, C, Q & D).	Mandatory directed training events that lead to specific certifications, qualifications, and or designations. Additionally, this phase provides Combat Leadership requirements.
7000	Reserved	Reserved for future use – to be assigned by ATD.

Phase	TERM	DEFINITION
8000	Academics	Training events to enhance professional understanding of Marine Aviation and the MAGTF. Includes position training for Aviation Ground communities and ACPM.
9000	Reserved	Reserved for M-SHARP use – to be assigned by ATD.

Table 1. The core model (after Headquarters United States Marine Corps, 2011)

The Core Skill Introduction phase is completed at the FRS. Core Skill and Mission Skill phases are completed at the operational squadron throughout the course of an aviator’s assignment to that unit. The Academics phase is continuous throughout and a supplement to each phase of training. The aviation career progression model (ACPM) is a series of academic presentations, readings and discussions that are meant to broaden Marine aviators’ knowledge and understanding of the operation of the Marine Air Ground Task Force (MAGTF). This phase is continuous throughout all phases and completion of certain ACPM events is a prerequisite to progressing into the next training phase. The core model is intended to integrate with the ATS and employ the concepts encompassed in the systems approach to training.

2. Readiness Reporting Tools

Several major readiness reporting tools are in use by Marine aviation. These include the Defense Readiness Reporting System (DRRS) Marine Corps, the Current Readiness Program (CRP), and the Marine Sierra Hotel Aviation Reporting Program (MSHARP). DRRS combines personnel and equipment levels with METs to inform upper echelons of command both at the operational and strategic levels. The CRP “is utilized by aviation commanders to maximize readiness, optimize resources (allocation and expenditures) and minimize logistical delays in order to produce core competent aviation units (squadrons/detachments),” (Headquarters United States Marine Corps, 2011, p. 7-3). The CRP utilizes metrics that measure the level of competency to which a unit is trained by aggregating information regarding the number of personnel trained to complete sub-sets of the METs trained to in the core model. Some information derived from the CRP is fed into DRRS. MSHARP is used at the tactical squadron level to

manage flight training plans, flight currency, and flight proficiency. It is important to note that minimum levels of both currency and proficiency are met merely by completion of events and flight hours, not the level of performance by which they are completed. The focus of this research is on the level of individual aviator performance and training, which can be aggregated to the battalion/squadron level for an understanding of personnel proficiency.

3. The Aviation Training Form and Grading Metrics

The current ATFs utilized by the AH and UH communities have evolved over time into their current form. An example can be seen in Appendix A. The form is standardized across the operational fleet for type and model of aircraft. Each event in the specific community *Training and Readiness Manual* has a corresponding ATF on which the PUI is rated using a criterion-based scale from zero to four. The grade of zero is assigned for any item that is graded as unsatisfactory. Unsatisfactory marks indicate "unsafe or complete lack of ability or knowledge," or "requires substantial input from IP for safe execution and/or mission accomplishment" (see ATF in Appendix A). The grades one through four correspond to the following criteria:

1. Safe but limited proficiency. Requires frequent input from the IP.
2. Correct. Recognizes and corrects errors. Requires occasional input from the IP.
3. Correct, efficient, skillful, and without hesitation. Requires minimal input from the IP.
4. Unusual high degree of ability. No further instruction required.

Instructors also have the opportunity to indicate that a particular item was not performed by selecting the did not do (DND) option. It should be noted that the ATFs for FRS events differ slightly in their criteria, and are enumerated as follows:

1. Consistently deviates from MDG standards. Slow to self-recognize errors with delayed or inappropriate corrections. Requires frequent IP coaching and/or control inputs to keep maneuver within safe parameters. Task saturated. Severely degraded crew resource management (CRM).
2. Deviates from MDG standards. Slow to self-recognize or requires moderate verbal coaching and minimal control inputs from IP for recognition and correction. Replacement aircrew (RAC) is working to actively employ CRM with lapses.

3. Autonomous with transitory deviations from MDG standards. PUI self recognizes and corrects in timely manner and/ or correctly self-debriefs. Situation appropriate CRM with minor lapses

4. Completely autonomous and within defined MDG performance standards. Situation appropriate CRM.

The differences between the two references are important to recognize if these values are to be used to analyze performance. For tactical squadron performance criteria focus on the amount of IP intervention, whereas in the FRS the focus is on compliance with the maneuver description guide, as well as level of IP input. There is no presumption of safe operation of the aircraft in the FRS, however, in the squadron all items are characterized as being completed safely.

The first section of each ATF is comprised of standard items: discussion items, brief/debrief, mission planning, checklists, communication, airwork, situational awareness, headwork, emergency procedures, and crew resource management. These items generally follow the definitions delineated by the Navy's CRM courses as well as those found in the *Naval Aviation Training and Readiness Program Manual.* The items most closely aligned and defined within CRM are communication, airwork, and situational awareness. Mission planning parallels the CRM principle of mission analysis. These items are defined in Table 2.

The Standard	***Below-Average/ Unsatisfactory Characteristics***
• ***SITUATIONAL AWARENESS (SA)***	
Demonstrate ongoing awareness of mission status and identify problems/potential problems and the need for action. *Maintain a proper scan pattern* *Monitor for trends, changes, and abnormal conditions, and share this information with other crewmembers* *Detect deviations from normal procedures and SOPs as well as task overload, underload, or tunnel vision of crewmembers* *Identify potential impact of problems to mission completion*	*Incomplete, sporadic, unaware, off track, or misjudged*
Clarify the validity of discrepant information (e.g., conflicting, ambiguous, incomplete).	*"Not my job," or unconcerned*
• ***ASSERTIVENESS (AS)***	
Ask questions when uncertain about decisions/procedures or objectives.	*Unconcerned, or too timid*
State opinions, advocate course of action, and make suggestions regarding decisions/ procedures. *Request information when needed; confront ambiguities and conflicts* *Make positive calls when safety of flight is threatened; declare an emergency when needed Offer/recommend alternative courses of action and/or mission alternatives; provide information without being asked*	*Apathetic, or intimidated*
• ***DECISION MAKING (DM)***	
Identify that a decision must be made based on situational assessment.	*Ignore the problem*

The Standard	***Below-Average/ Unsatisfactory Characteristics***
Gather, crosscheck, and evaluate information sources (other crewmembers, ATC, metro, headquarters, support, instruments/equipment) prior to making a decision; filter out erroneous/irrelevant information.	*Jump to conclusions; be misled by poor information*
Generate and discuss alternatives using relevant data; provide rationale for all decision alternatives.	*Bias, "My way or else," close-mindedness*
Anticipate the consequences of a decision alternative.	*Not thinking things through*
Choose the best alternative, communicate internally and externally, and evaluate its effectiveness.	*Indecisiveness, rigidity, faulty communications*
• ***COMMUNICATION (CM)***	
Provide appropriate response to a communication (e.g., acknowledge, repeat, and request clarification).	*Ignore, respond to the feeling, incorrect response*
Use standard terminology and non-verbal signals with accurate, timely, and concise information.	*Inefficient, vague, off the subject*
• ***LEADERSHIP (LD)***	
Direct and coordinate the activities of other crewmembers; delegate tasks to other crewmembers.	*Ignore others, disregard*
Monitor other crewmembers to see if they understand what is expected of them; maintain constructive atmosphere.	*Discount others, selfishness, hostility*
Encourage crewmember participation; provide constructive feedback to other crewmembers.	*Disregard, prejudice*
• ***ADAPTABILITY/FLEXIBILITY (AF)***	
Alter plans and behaviors to meet situation demands; continue to function during system failures/malfunctions/changed mission.	*Inflexible, sudden loss of judgment, tunnel vision*
Step in and help other crewmembers; be receptive to input from other crewmembers. Adapt to personality styles of other crewmembers Accommodate and cope with stress of other crewmembers and self	*Lack of empathy, rigid, prejudiced*
• ***MISSION ANALYSIS (MA)***	
Conduct thorough pre-mission planning and briefings, assembling mission information, estimating mission timing, and setting priorities based on mission requirements.	*Haphazard, incomplete, mistakes, inattentive*
Devise contingency plans for unplanned events.	*Unprepared, no backup plans*
Report ongoing challenges to the mission plan; offer alternatives.	*Apathetic, no backup plans, intimidated*
Conduct thorough post-mission debriefs, effectively using feedback techniques.	*Incomplete, errors, omissions*

Table 2. CRM principles, definitions, and descriptions of acceptable and unacceptable performance (from Headquarters United States Marine Corps, 2011, pp. E-4 - E-5).

It should be noted that not all of the items are addressed by the principles of CRM. With the exception of headwork and CRM itself, there are no formal definitions for the items, but they are self-explanatory when coupled with the criteria outlined within the ATF. Headwork is formally defined in the *Student Naval Aviator Training and Administration Manual*.

> Headwork is the ability to understand and grasp the meaning of instructions, demonstrations, and explanations; the faculty of remembering instructions from event to event; the ability to plan a series or sequence of maneuvers or actions; the ability to anticipate and avoid possible difficulties; and the ability to plan and execute alternative options. (Naval Air Training Command, 2007, p. VII-4)

This definition closely aligns with the definition of decision making found in Table 2.

According to COMNAVAIRFORINST 1542.7 CRM is defined as follows: “The effective use of all available resources by individuals, crews and teams to safely and efficiently accomplish the mission or task. CRM also refers to identifying and managing the conditions that lead to error” (Naval Aviation Schools Command, 2013).

Each ATF also has mission-specific items that are evaluated by the instructor based on the requirements for the training event. In addition to a numerical grade instructors may provide remarks for each item, and are afforded an opportunity to provide additional comments in a free text box. A numerical average of all items that have received a mark is calculated and recorded on the ATF. The instructor pilot (IP) also marks whether the training event is unsatisfactory, complete, incomplete, or the PUI requires additional training. Unsatisfactory flights are considered derogatory and reflect poorly on a PUI’s record. A completed flight indicates that the PUI has met all the requirements in the training and readiness manual, and the IP is satisfied with the PUI’s performance. It also indicates that the PUI is ready to move on to the next event in his or her respective syllabus. Incomplete flights indicate that the training and readiness requirements could not be met due to weather, aircraft maintenance or other unforeseen, limiting circumstances. If a PUI receives the grade of Requires Additional Training (RAT), it is not considered derogatory towards his or her performance record and indicates the PUI needs greater time and exposure to certain maneuvers or concepts in the IP’s opinion.

In the case of either an unsatisfactory or RAT event, the IP is responsible for developing a course of action to remediate the PUI. The development of that course of action is to be endorsed by the squadron leadership, and seen through to completion by

both the instructor and the command. The remediation plan is to be adhered to by the PUI, who will get another opportunity to attempt the event. Assignment of these grades to PUIs are rare and are considered gravely by training staff before being assigned for a number of reasons that include the impact on required resources, overall training progression of the individual, possible stigma associated with receiving either of these grades, and the overall readiness of the squadron.

D. INSTRUCTIONAL DESIGN

To begin understanding the process of assessing student performance, one must analyze and evaluate the concepts underlying instructional design, and the context in which instruction is taking place. An understanding of learning processes and theories is also necessary to implement a system of instruction that is effective. Gagné and Briggs's (1979) model of instructional design specifies 14 stages (p. 23). Among these stages, those that are important here are the sixth stage, "Definition of Performance Objectives," and the ninth stage, "Assessing Student Performance (Performance Measures)" (p. 23). In defining performance objectives, designers must develop a strategy to specify how broad (or narrow) they intend for the specified objectives to be (Gagné & Briggs, 1979, p. 31). Regardless of how broad or narrow objectives are defined, they should be defined as precisely as possible. This precision allows learners to understand not necessarily how they will achieve success, but rather, how they could observe it themselves (Gagné & Briggs, 1979, p. 119). The assessment stage requires that designers specify what method or combination of methods they will utilize to evaluate student progress. In addition, they must ensure that whatever methods are chosen are in concert with achieving the performance objectives for the course of instruction. We can infer from this that student performance assessment can only be conducted within the frame of reference provided by the structure and design of the instructional system.

The analyze, design, develop, implement, and evaluate (ADDIE) model for instructional design is also a useful tool (see Figure 1). This thesis focuses on the portion of the model that addresses primarily the analysis, design, and evaluation stages. Branch (2009) identifies a number of methods to carry out this evaluation to include surveys,

observations, and supervisor reviews (p. 160). Since the instructional system for naval aviation exists, we can begin at the evaluation stage and, then conduct analysis, and improve the current design.

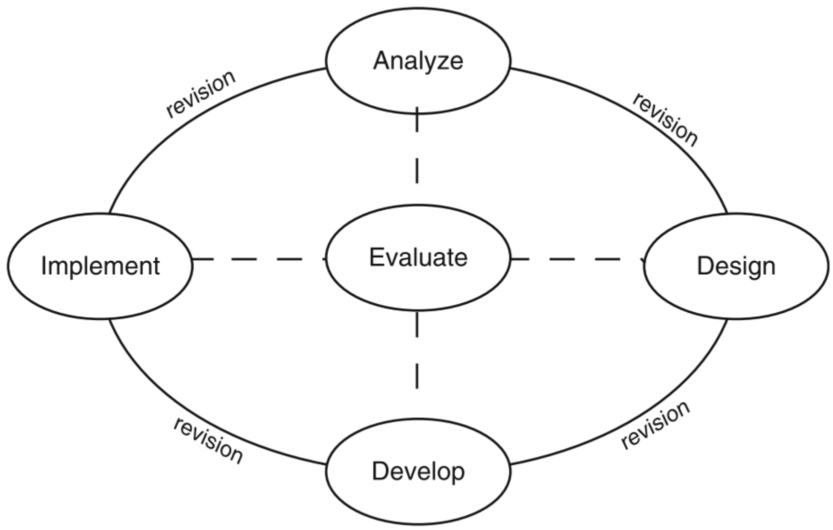


Figure 1. The ADDIE framework (from Branch, 2009)

1. Learning Processes

There are a number of models that attempt to define the intent for instructional systems. Gagné and Briggs (1979) describe five specific learning outcomes that underlie the intent behind an instructional system: intellectual skills, cognitive strategies, verbal information, motor skills, and attitudes (pp. 49–50). Intellectual skills can be defined as the comprehension of underlying concepts. The mere ability to recite the existence of some facts does not qualify as intellectual skill (p. 49). The ability to synthesize information from these facts and be able to apply the knowledge of these facts in the appropriate situation would qualify as intellectual skill. In the context of aviation instruction, an example might be a pilot under instruction understanding that the process of flight planning requires that they evaluate the forecasted weather. Simply addressing whether the minimum visibility and cloud ceiling requirements are met is not intellectual skill. Instead, intellectual skill involves, for example, recognizing that even though the minimums to fly the aircraft are met, the ability to employ weapon systems may still be

in question. A cognitive strategy is the internal method that a learner uses to solve problems. Once learners adopt a strategy, they may call on it in the future when faced with similar problems (Gagné & Briggs, 1979, p. 50). Verbal information is knowledge, such as the days of the week, or historical facts that are recalled often and remain in a person's memory over the course of a lifetime and can be recalled when required (p. 50). Motor skills are self-explanatory, and have a clear correlation to the aviation training domain. Finally, the fifth learning outcome is developing attitudes. Attitudes, as defined by Gagné and Briggs (1979), "amplify an individual's positive or negative reactions toward some person, or thing or situation," (p. 50). These five learning outcomes comprise the "capabilities of human performance," (Gagné & Briggs, p. 51). The collection of these capabilities encompasses the performance ability of an individual, but more importantly, break down the meta-performance into sub-categories that are more easily measurable.

Examining attitude learning in greater detail, direct and indirect methods exist, of which, both are used in naval aviation training. The direct method is at its base, reinforcement learning. An example of the child touching a hot stove and not repeating that behavior would be a type of negative reinforcement. Positive reinforcement can also occur, as in an example of providing some benefit after the student or trainee exhibits a desired behavior. In contrast to the direct method is the indirect method, which focuses on human modeling (Gagné & Briggs, 1979, p. 88). In this case a learner observes attitudes and behaviors, and in some way respects, or identifies with the individual displaying the attitude or behavior and is led to mimic his or her observations (Gagné & Briggs, 1979, p. 89). Human modeling plays a significant role for Marine or naval aviators under instruction as they are aspiring to achieve the qualifications and designations that those that are instructing them hold.

2. Mastery and Diagnostics

Within the realm of any educational pursuit mastery of the concepts and skills that are being taught is the ultimate goal. Mastery is achieved based on several factors: aptitude, quality of instruction, the ability of the student to understand the instruction,

perseverance of the student, and the time allotted for learning (Bloom, Hastings, & Madaus, 1971). This is a relatively long list of factors each of which has a considerable amount of variation among differing environments. Bloom et al. (1971) assert that use of the Normal curve is not sufficiently representative of student performance (p. 45) and the expectations of instructors play a significant role in student achievement. If a teacher expects one third of their class to fail or barely pass, one third to be considered simply "just ok," and one third to be capable, with even a smaller percentage excelling, then this will be the case, especially when coupled with the use of the Normal curve for grading. In regards to Marine aviation mastery is sought out from the earliest stages of flight instruction; however, there is little expectation of mastery of skills early on. There is an incremental approach to building basic skills and then compounding those skills with new requirements. At the tactical squadron, the initial expectation of PUI is that they are capable co-pilots and aviators who can fly the aircraft in a safe manner in both day and night environments. It should be noted that in contrast to the one-third distribution previously discussed, Marine aviators are expected to be capable of achieving mastery as they progress through the course of instruction. However, there is no data to support what expectation is held by instructors of PUIs within the tactical squadron. Therefore, for purposes of demonstration only, if IPs expect one in five PUIs to be incapable, and three in five to be capable, and one in five to excel, this may be the outcome. While instructors and commanders are determining the level of mastery of the PUI, the results may match this distribution as a matter of expectation of instructor expectation rather than PUI capability.

Diagnostic evaluation serves to assign value, determine, describe, and classify student behaviors in some way (Bloom et al., 1971). Diagnosis can be performed at different times during the course of instruction, including pre-instruction. If done prior to beginning the course of instruction it is intended to ensure that the student or trainee possesses the prerequisite knowledge to proceed, and would be considered a summative assessment. Conducting diagnosis mid-course of instruction intends to address repetitive shortfalls in student learning of specific concepts or skills. Mid-syllabus diagnosis would

be considered a formative assessment; however, generally, diagnosis serves a primarily summative function (Bloom et al., 1971). Summative and formative assessments are defined and discussed in depth in section E.1.

3. Program Evaluation

In attempting to investigate the method by which trainees or students are assessed one must also consider the entirety of the instructional program. This becomes of particular interest in fields that require training of specialists that are required to develop in-depth technical knowledge that supports subjective decision making skills in an infinite number of scenario permutations, no two of which are exactly alike. Training military aviators certainly fits this description, as does the training of medical doctors. Both of these fields have unique technical aspects that are taught through a combination of classroom-based and practical experience-based instruction. Musick (2006) provides a discussion utilizing a similar conceptual model to that offered by Gagné and Briggs (1979) regarding instructional design. Figure 2 summarizes his conceptual approach.

Task-Oriented Conceptual Model of Program Evaluation in Graduate Medical Education

Step One: Determine evaluation need

WHY is the evaluation being undertaken and for whom? (Accreditation requirement; institutional requirement; specific project; research)

Step Two: Determine evaluation focus

WHAT entity is to be evaluated? (Overall training program, clinical rotation, didactic event, teaching faculty, residents/fellows)

Step Three: Determine evaluation methodology

WHEN is the evaluation procedure to be undertaken? (Planned clinical observation, end of rotation, end of year, after graduation)

WHERE are evaluation data to be collected? (Normal patient care settings, classrooms, other)

HOW are evaluation data to be collected? (Ratings of performance, written/oral examinations, attendance sheets, rotation objectives checklists, surveys, clinical skill examinations)

WHAT types of data analyses will be needed? (Reporting formats, data properties/psychometrics)

Step Four: Present evaluation results

WHO are the key stakeholders who must review the results? (Department chair, teaching faculty, institutional GME personnel, residents)

WHEN should results be presented? (Regular agenda item for faculty meetings; annual program evaluation meeting and/or educational retreat; education committee meetings)

Step Five: Document evaluation results

HOW are evaluation results documented and used for program improvement? (Content delivery issues, frequency with which outcomes are measured, program changes made as a result of evaluation data, resident input into program improvements)

Figure 2. Task-oriented conceptual model of program evaluation in graduate medical education (from Musick, 2006, p. 800)

There is a clear need to ensure that doctors, once complete with their graduate medical education, have learned the requisite knowledge and skills to carry out their duties as a medical doctor acting under their own recognizance. Aviation training has similar requirements. Once aviators complete a course of instruction and their commanding officer designates them as qualified to perform certain types of operations, the expectation is that they will capably manage their aircraft in the applicable situation. Musick (2006) notes the emphasis on an outcome-based approach of program evaluation versus a process-based approach in the medical community with respect to program evaluation (p. 759). A process-based approach is an evaluation methodology in which only the completeness and organization of the system or curriculum is examined. In contrast, outcome-based approaches consider not only the thoroughness of the system or curriculum, but also consider the attendance and performance of those trainees within the

system or curriculum. An example of a process-based approach is the accreditation process for a university major study program that evaluates only whether the syllabi that are offered to confer the degree on graduates are thorough enough in the discipline to warrant the issuance of the degree. This same example would become an outcome-based approach if the program was also evaluated on student attendance, performance on a standardized test, and perhaps even the percentage of students who are able to find work in the field upon receiving the degree. The outcome-based approach most closely models the current Marine and naval aviation model, where the de facto accreditation is for a unit to have the appropriate number of pilots qualified to carry out a number of different skill sets as delineated in the *Training and Readiness Manual*. The key point borne out by Musick (2006) is that the entirety of the instructional system must be taken into consideration and to truly evaluate the effectiveness of graduate medical education substantial effort must be made to design a comprehensive system of instruction that effectively measures the outcomes that have been determined to be acceptable within their domain.

E. EVALUATION METHODOLOGIES

The primary means by which Marine aviators are assessed is by observation of a PUI's performance by instructors during execution of training events enumerated in their community's training and readiness manual. This observation is recorded on an ATF filled out by the instructor. To understand evaluation we must discuss what can be evaluated in instructional systems, how these evaluations are constructed and what they measure.

1. Summative and Formative Assessments

Summative and formative assessments are inextricably linked; however, each has its own distinct assessment purpose. To begin to analyze how these two assessments are related, we must examine the definition of each. Summative evaluation is concerned with the general level of understanding of a concept or concepts over the full course of instruction or a large portion of the course (Bloom, Hastings, and Madaus 1971, p. 60). In contrast, formative assessment is intended to "determine the degree of mastery of a given

learning task and to pinpoint the part of the task not mastered" (Bloom et al., 1971, p. 62). Summative evaluations are conducted with less frequency than formative assessments. Harlen and James (1997) point out that formative assessment is intended to provide feedback for the instructor and the learner about current levels of understanding and how to formulate the future course of instruction (p. 369). The course of instruction following a formative assessment is then developed to allow the learner to make strides towards mastery of the subject, skill, or concept.

Some educational researchers argue that the summative and formative evaluations have become confused in modern educational processes (Harlen & James, 1997), and some allude to the demonizing of summative evaluations due to the implications of eliciting a judgment of learner performance (Taras, 2005). Taras (2005) argues that the two forms of assessment, which in some cases are placed in a rival role, should be complementary (Taras, 2005). This is in concert with the conceptual model enumerated by Bloom and others (Bloom et al., 1971), and suggests that a balanced and blended approach between the two methods be utilized. The intended use of formative evaluation is to continually provide the learner with periodic updates to the level of mastery, and gaps that exist within the knowledge obtained, while summative assessments are intended to provide a broad, generalized summary of student capability within the subject matter.

2. Criterion Referenced Performance Assessment

Criterion-referenced measurements are designed to evaluate the abilities of a person to complete specific tasks based on what has been operationally defined, and capable of being both observed and measured. Swezey (1981) points out that despite the wide-spread use and acceptance of norm-referenced measurement, it may not always be the most appropriate method by which to evaluate a learner or trainee (p. 5). The major difference between norm-referenced and criterion-referenced measurements is that norm-referenced measurements compare performances of individuals with that of a particular group, while criterion-referenced measurements compare individual performance to a well-defined set of operationally contextual standards. According to Swezey (1981), criterion-referenced measurement can include either domain-referenced or objectives-

referenced measurement models, or both (p. 8). Domain-referenced measurements focus on eliciting information from groupings of items that are representative of all potential test items, and objectives-referenced evaluations focus on the targeting of specific behaviors. The primary difference between these methods and criterion-referenced measurement is that they are focused on the content of testing, rather than interpreting the scores elicited in evaluation (Swezey, 1981, p. 7).

The focus on the interpretation of results should mirror the evaluation methods of Marine aviators undergoing a particular training syllabus, especially given that "criterion-referenced measurement . . . is usually the measurement model of choice when judgments are desired about an individual's achievement of specific objectives," (Swezey, 1981, p. 11) . In the model proposed by Swezey (1981) for criterion-referenced measurement, he enumerates three separate characteristics of criterion-referenced measurements: test scoring based on absolute standards, a primary focus on measuring a level of mastery, and known performance objectives associated with a task (p. 10). Criterion-referenced tests may be used for multiple reasons; however, in regard to aviation training and developing a training diagnostic tool, we are primarily concerned with using them as an aid to diagnosis of a PUI's performance, and as a tool for evaluating the instruction received. Swezey (1981) proposes seven steps to developing the criterion-referenced test, of which we will focus on evaluating input to the development process, planning the test, and test administration and scoring.

In the evaluation phase the most critical activity is conducting an in-depth task analysis that addresses the requisite skills and knowledge, necessary performance of a subject, identifies the specific criteria correlated with each performance, and identifies the conditions under which the performance is required to be completed (Swezey, 1981, pp. 23-24). This is critically tied to the development of objectives. In order to develop effective objectives, we must be specific in their intent, ensure that the scope of the objective is narrow enough to be measured, and use precise operational language (Swezey, 1981, p. 24). By decomposing objectives into three component parts, performances, conditions, and standards (Swezey, 1981, p. 25), the criterion-referenced measurement developer can effectively construct methods that are effective in collecting

the information desired. The primary goal of developing objectives, then, is to ensure they are unambiguous, specific enough in the domain, and their intent is clear.

The planning of a criterion-referenced test requires the author(s) of the test to ensure that they take into consideration all of the constraints and restraints that might have an effect on the implementation of the test. Swezey (1981) provides a short list, of some of the more common practical constraints, which include testing time available, weather conditions, geographic limitations, personnel limitations, equipment available, realism, and cost limitations (p. 46). All of these constraints play a role in the management of a military aviation training syllabus.

Despite criterion-referenced measurements often being used for “pass-fail” type evaluations, this is not a limitation. In addition, one might argue that despite the grading scale currently found on ATFs and enumerated in the T&R manual that current practice actually equates to a pass-fail system. The intent for it to be a graduated scale that allows for instructors to discriminate between the performance of individuals versus only knowing which trainees are qualified and which are not is lost in the failure to effectively apply empirical analysis. Swezey (1981) addresses rating scales by recommending behaviorally-anchored rating scales because they provide the strict definitions required for the rating scale (p. 64). Because these types of scales require judgments to be made by the rater, they can be susceptible to a number of different errors. Swezey (1981) describes four categories of rating error: error of standards, error of halo, logical of error, and error of central tendency (p. 66). The first of these errors results when the standards are not adequately described. The error of halo results when the rater forms an impression of the person being rated, either positive or negative, and biases their ratings in that direction on the rating scales. Logical error results when the rater makes an erroneous correlation between two distinct behaviors that are independent of one another and rates both items/behaviors in a similar fashion. This can be a common mistake of instructor pilots and is specifically addressed by the *Training and Readiness Manual* in regard to the items of “Headwork” and “Situational Awareness.” Finally, the error of central tendency

is the predisposition of raters to force their scoring to mirror the normal curve, with most students being rated as middle performers, and fewer that are high and low performers, respectively (Swezey, 1981, pp.66–67).

3. Behaviorally Anchored Rating Scales

The development of a Tactical Thinking Behaviorally Anchored Rating Scale (T-BARS) was undertaken by the Army Research Institute for the behavioral and social sciences in pursuit of an assessment tool to measure the tactical cognitive skills of officers in the combat arms (Phillips, Shafer, Ross, & Cox, 2006, p. 2). This research has direct application to the assessment of Marine aviators in their respective tactical squadrons. Although a rating system already exists within each model of aircraft's training and readiness manual, the T-BARS provides a frame of reference on how to interpret the existing rating system in the aviation community. The T-BARS methodology also suggests that the development of the system in the aviation community may be incomplete. The development of the T-BARS by Phillips et al. (2006) utilizes the Dreyfus and Dreyfus (1980) five-stage model of skill acquisition. One of the critical components of the research was the establishment of inter-rater reliability when evaluating the application of the T-BARS (Phillips et al., 2006, p. 21). In Marine aviation, there currently are not any inter-rater reliability measures among instructor pilots. This is a point for further investigation and discussion while attempting to characterize PUI performance. Finally, the authors postulate that the T-BARS be used "in order to determine the optimal course of instruction to develop him or her into a well-rounded tactical thinker" (Phillips et al., 2006, p. 24). By utilizing a similar methodology, the data contained within the aviation tracking files of a pilot's training record can potentially provide similar details for informed training interventions.

4. Debriefing As Part of Assessment

The practice of using debriefing to enhance learning, and formulate new methods to approaching tasks, has been widely used in the military for many years. Within Marine aviation, the accepted method for debriefing within the H-1 community is the *NTTP 3-*

22.5 Element Debrief Guide. The element debrief in Figure 3 guide outlines 15 items to discuss during the post-mission debrief and provides a model to discuss all aspects of the flight.

ELEMENT DEBRIEF GUIDE

- SOF & ALIBIS
- ESTIMATE OF MISSION SUCCESS
    - MISSION SUCCES CRITERIA
        - REVIEW
    - WEAPONS RELEASE EVENTS
        - EFFECTS OF FIRE
    - SURVIVABILITY
        - THREAT / TACTICS
- OBSTACLES TO EFFECTIVENESS
    - DEVIATIONS FROM SOP
        - UNBRIEFED
    - PLAN / BRIEF / EXECUTION
    - LESSONS LEARNED
        - THREE GOOD / BAD
    - OTHERS

Figure 3. Element Debrief Guide (from Naval Air Systems Command, 2011, p. 94)

This debrief is considered a part of the actual flight event itself and is a critical part of the training process. It is usually led by the pilot under instruction, moderated by the lead instructor pilot of the event, which always involve multiple aircraft crews. Debriefs can provide an environment conducive to formative assessment, which has also been acknowledged by the medical community (Rudolph, Simon, Raemer, & Eppich, 2008). Rudolph et al. (2008) offer a four-step model for debriefing as formative assessment. They point out that “the hidden curriculum of assessment includes implicit feedback about how well the trainee is performing a new professional role” (Rudolph et al., p. 1011). This certainly applies to aviators under instruction as well. The four steps outlined by Rudolph et al. are first to note the gaps in performance from those outlined by objectives, second, to provide feedback describing the shortcomings to the learner, third, examine why the gap exists, and fourth, fill the gap through the relevant guided discussion and instruction (Rudolph et al., 2008, p. 1010) . This model accurately describes the intent of the element debrief when used for the purpose of debriefing a training flight. Rudolph, et al. (2008) also describe the usage of debriefing as a formative

assessment in depth, by specifying that first the context for learning is defined, that objectives are provided and effective by being observable, and the debriefing provides phases for the learner's reaction to the event, analysis of the event, and summary of the event (Rudolph et al., 2008, pp. 1012–1013). The usage of the element debrief when viewed through the lens of this model provides feedback to both the learner and the instructor. The instructor is able to assess the trainee's perception of the event and whether gaps exist. The knowledge gained by the trainee is formative in the sense that it provides an opportunity for the learner to self-identify existing gaps. In this manner all participants of the debrief are able to identify strengths and weaknesses in the problem-solving approach used in the scenario.

5. Evaluation in Military Aviation

Based on the previous discussion of summative and formative assessment, mastery learning, and diagnosis it becomes apparent that within military aviation, both summative and formative evaluations are conducted simultaneously. As PUIs progress through each event in the course of instruction they are evaluated on a number of different skills and concepts. Some of these skills and concepts such as "air work" and "situational awareness" are repeated throughout the course of instruction, while others are specific to a particular training event. PUIs are expected to learn new skills throughout the course of instruction in order to enhance their ability to conduct any mission the unit has a potential to be assigned. Summative assessment is provided in the form of a numerical grade following each flight, which is simply an average of the numerical score on each assessed item for the particular flight event. The most relevant form of summative assessment is the Navy Standard Score (NSS), which is calculated using descriptive statistics and norming methods (Naval Air Training Command, 2007, Appendix E). The NSS is utilized only until an aviator is designated as such. Formative assessment is provided each flight as well though the same vehicle of comments and numerical grade assigned by the instructor. Formative assessment is also provided via a flight debrief with the crew following every event. All of the assessments conducted are utilized for diagnosis by instructors and leadership to evaluate a PUI's ability in the cockpit and the ability to progress in the syllabus. Harlen and James (1997) point out that

"summative assessment should mean summing up the evidence, not summing across a series of judgments or completed assessments . . ." (p. 375). This is precisely what is occurring in the assessment of PUIs in Marine aviation. We do not suggest that the current assessment methods are inappropriate in their existence, but rather improperly interpreted and underutilized with respect to the information available. Furthermore the current formative assessment provides fractured and imprecise feedback to the trainee.

F. DECISION SUPPORT SYSTEMS AND DASHBOARDS

Managerial decision making can be a complex, high-stakes process, and nowhere is this truer than in the military service. With the advent of more robust technology and computational power more data is being collected than ever before. Despite the vast amounts of data being collected across a multitude of domains, there remains the need to reduce the data to a manageable size and enhancing its meaning to those that are interested in it and supporting managerial decisions. A common method for approaching this problem is the utilization of dashboard applications. The evolution of computing power in the 1970s laid the ground work for dashboard applications as decision tools for management information systems, executive information systems, and decision support systems (Beuschel, 2008). Many of these systems focus on business decisions regarding how companies can increase their bottom line, by appealing to customers more efficiently, or comparing a number of potential outcomes of different decisions. Breuschel (2008) states that decision support systems "address decision problems where the solution-finding process may not be completely structured" (p. 116). In the case of Marine aviation, senior members of the squadron staff as well as the instructor cadre must understand how the pilots undergoing a training syllabus are performing. The current manner in which this knowledge is obtained is through review of training forms and through discussions among instructors about individual and group performance, which is hardly a structured problem space. A distinction must be made here between management information systems and decision support systems. Management information systems simply summarize and provide reports on basic operations of an

organization, enterprise or institution, while decision support systems are focused on addressing the problem-space by bringing to light information that makes solutions more apparent (Breuschel, 2008, p. 116).

There are several existing models of how decision support systems should be developed and what they include. One of these models referenced by Brueschel (2008) states that when in the form of a dashboard, the decision support system includes three components, which are visualization, relevant data selection, and monitoring and interaction (p. 117). Visualization might be considered as the most important of the three components because it is a tangible factor; however, it is of equal importance as both relevant data selection and monitoring and interaction. Visualization offers users their initial glimpse of the program, data, and information that the system has to offer. If it is difficult to discern what information is being presented in a visualization, then regardless of the data presented or the level of interaction, the system becomes less usable. The most simplistic example of visualization for decision making is the “stop-light” chart, which provides the user with red, yellow, or green cells highlighted to indicate unacceptable, at risk or, acceptable, respectively. In regard to all decision support systems with respect to visualization the intent is to “indicate a potential need for action” (Beuschel, 2008, p. 117). Selecting relevant data must also be considered, and seems to be an obvious component. One would expect that a decision support dashboard would utilize the necessary and pertinent information to the domain; however, data must be summarized and compressed, which results in the loss of some granularity. Finally, the third component of monitoring and interaction also must be given equal consideration. It is this component that must be properly designed to allow users to achieve the granularity of data necessary to inform their decisions, as well as ensure that the data that is provided to decision makers is current.

Averweg (2008) addresses the issues of decision making in categorical terms: independent, sequential independent, and pooled interdependent (p. 218). Averweg (2008) states that the primary value of the decision support tool is to allow the exploration of the data available by the user to provide the ability to identify and compare several courses of action. One of the challenges in designing a decision support tool, with

regards to Marine Aviation, is that while the commanding officer has the final say, he generally takes into account the trusted input of the instructor cadre and senior staff. These multiple individual perspectives could make the distinction of key pieces of information opaque, when there are starkly conflicting opinions. Paradice and Davis (2008) offer a model that attempts to address the conflicting perspectives by viewing them as either technical, personal, or organizational. To summarize, they believe that when the decision support system is designed it must take into account each of these categories, in order to be implemented in such a manner that is useful in the domain for which it was designed. This remains important as we must manipulate the data in the system in the background in order to compress it into relevant and understandable snapshots. Decision makers must understand who is providing them with data, that the data being presented to them is relevant to their cause, and finally, that the data meets the organizational intent of the institution.

1. Models for Decisions Support Systems

There are a number of existing models that attempt to provide a framework for managerial decision making. These models are utilized to help develop decision support systems in a vast array of different domains. These domains range from learning and efficacy to best business practices to medical treatments. In Marine aviation, the decisions we would like to support revolve around how to have individual aviators progress through their training syllabus and how to focus instructional efforts to meet the needs of trainees. While the models researched are not directly related to training or aviation, they have potential to be adapted to support the decision-making processes of squadron leadership.

The first of the models studied is the classification and ranking belief simplex (CaRBS), developed by Beynon (2005), which attempts rank and classify objects to a specific state. In Benyon's methodology, objects are defined by measurements of a collection of variables that support either a hypothesis or its complement. The CaBRS utilizes the Dempster-Shafer theory of evidence as a foundation that provides for the allowance of uncertainty within the data set by modeling the "presence of missing

values" (Beynon, 2005, p. 76). This model could be considered useful in evaluating ATF data by classifying each of the standard graded items as objects and utilizing their score values as their level measurements. The benefit of this model is that it allows for "ignorant values" (Beynon, 2008a). CaRBS produces, as graphical output, points within a triangle whose vertices are the hypothesis, its complement, and the set containing both, indicating uncertainty.

The challenge in applying this model in the aviation context is two-fold. First, the hypothesis and its complement must be considered. This suggests that the hypothesis would be that the particular PUI is capable at necessary tasks and its complement. Second, there is the usage of the one through four grading scale of discrete values as low-level measurements. The values are treated as continuous when averaged on the ATF despite being discrete criterion references. This might be resolved by considering the trainees as objects and taking their overall average scores on each flight as the individual measurements to consider for ranking. This process could allow instructors to make comparisons among those progressing through a training syllabus of whom to accelerate and who needs additional attention; however it does not offer insights into the areas in which training intervention is needed.

A second method that may be of particular use is qualitative comparative analysis (QCA). Beynon (2008b) states that "QCA is employed in comparative case-oriented research, for studying a small-to-moderate number of cases in which a specific outcome has occurred, compared with those where it has not" (p. 751). However, this method differs from many typical statistical comparison methods that rely on the evaluation of independent variables individually. QCA relies on differing combinations of variables and comparing their effect on independent variables. It does this by using Boolean algebra to make comparisons for each case combination (Beynon, 2008b, p. 751). The QCA discussed by Beynon (2008b) relies on the Quine-McClusky method to reduce the equations entered into the truth table (p.752). A limitation of QCA is that too many variables may obscure the underlying implications of the data (Beynon, 2008b, p.754). Finally, Beynon (2008b) states, "QCA is associated with policy based decision making, where a common goal is to make decisive interventions," (p. 754). This is a

representative statement of the goal of squadron leadership, where policy decisions refer to the administration of the training programs for groups and individuals as well as the management of the instructor cadre.

The other two models investigated for this research by Power (2008) and (Beynon, 2008c), after further investigation, did not provide models that would be relevant to the decision-making needs of the aviation community discussed in this paper. Power's (2008) suggests real options reasoning; however this model is most well suited for business and financial market application. Although it does provide some insight into situations with uncertainty and how those decisions must be made with regard to acceptance of potential risk, it has limited applicability for decisions regarding how to train individuals. The PROMETHEE (preference ranking organization method for enrichment evaluation) is similar to the CaRBS; however it uses pairwise comparisons between values describing the alternatives (Beynon, 2008c, p. 743). This particular method also does not provide graphical representation of results.

2. Design of User Interfaces

The usage of computer systems to provide ease of access to information and simple and intuitive manipulation of information and analysis of data has become commonplace. Whether it is business analytics or medical applications, the advancement of computing power has made the use of these tools very popular. How users interact with a system is a critical component of their ability to interpret the information they provide. The field of human computer interaction has given rise to the term user experience, which generally refers to both practical and aesthetic factors of usability of a program or system over its full life-cycle (Hassenzahl & Tractinsky, 2006). When contemplating a computer-based decision support tool, we must investigate the topic of user experience so that the resulting tool not only supports the end-users' goals, but has a degree of user satisfaction that increases the users' desires to make use of the tool. Forlizzi and Battarbee (2004) state, "The term 'user experience' is associated with a wide range of meanings, and no cohesive theory of experience exists for the design community," (p. 261). They further argue that as a result of the lack of a well-defined

conceptual model or definition, user experience is a wildly diverse and striated field. Despite the topic of user experience being broad, we will utilize the summarization provided by Hassenzahl and Tractinsky (2006), which is shown in Figure 4.

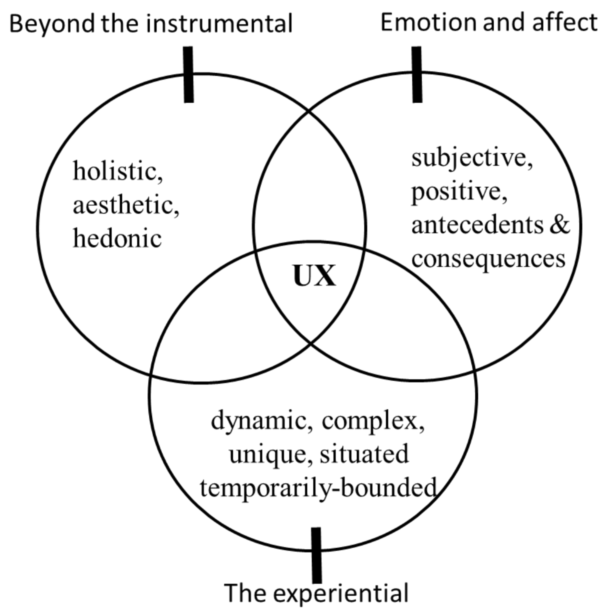


Figure 4. Facets of user experience (from Hassenzahl & Tractinsky, 2006, p. 95)

Some have categorized user experience models into three separate subcategories: product-centered models, user-centered models, and interaction-centered models (Forlizzi & Battarbee, 2004, p. 262). While much focus has been on the product-centered models, there has been a shift towards user- and interaction-centered models to understand the user experience (Hassenzahl & Tractinsky, 2006). Regardless of the type of model offered the intent is to support design to ensure achievement of the appropriate user experience. Forlizzi and Battarbee (2004) offer an interaction-centered model that possesses two subcategories, namely, the type of interaction, and the experience that results from the interaction (p. 263). The key concept in relation to the design of an interactive system is that the interactions must be palatable to the user on a fluent, cognitive and expressive level (Forlizzi & Battarbee, 2004, p. 262) and the experiences had during these interactions must also be positive.

A slightly different perspective on user experience is offered by Sutcliffe (2010), who distinguishes between user experience and, what he terms "user engagement" (p. 1). Sutcliffe's definition (2010) of user engagement "has a more restricted sense" than user experience that "focuses on the quality of the interactive experience rather than the whole life span experience of a product" (p. 1). Our focus is designing a decision support tool for training intervention that results in a positive user experience with the outcomes and trust of the system and its operation. For example, if an instructor sees such a tool as a 'black box' that simply provides information and he or she does not comprehend how that information is derived, he or she will likely judge the tool as unreliable. Sutcliffe (2010) believes that aesthetics may play an initial role in engagement; however, decisions and judgments are refined through continued use (p. 6). With respect to the aviation training domain, Sutcliffe (2010) points out that professional or "work domains involve slow path-decisions and usability/utility criteria" (p.6). This certainly is intuitive for decisions that require careful reflection and may have long-term impacts on the development and career progression of trainees. Another critical point made by Sutcliffe (2010) is that negative experiences tend to have a larger effect on users than positive ones. If users experience frustration or difficulty they will discount the product and be less inclined to utilize or seek experiences in the future to use it (Sutcliffe, 2010, p. 7). Rassmussen (1986) provides a framework that Sutcliffe (2010) states to be a useful model when addressing user engagement termed the Knowledge-Rules-Skills model. In the model, rules are the instinctual usage of the product, knowledge is a higher level, and skills are what support the building of new understanding of the product operation. Another assertion by Sutcliffe (2010) is that in "work/goal oriented applications, skilled operation and efficiency will be more important; hence, ease of learning and ease of use are paramount" (p. 8). This is the aim of developing a tool for instructors and squadron leadership to aid in training diagnostics. Such a tool would require efficient and intuitive use so that the intended audience will use it often enough to have an impact.

Finally, Sutcliffe (2010) offers three typical methods for the user engagement design process: the use of scenarios, the use of storyboards, and the use of personae (pp. 17–18). Regardless of which of these three methods is undertaken, he further offers a list

of principles that he recommends should be considered: immersion and presence, flow and interaction, media for mood and arousal, media to attract and persuade, media for emotional effects, media to attract attention, and design for aesthetic appeal (Sutcliffe, 2010, pp. 25–28). While all of these are important, some have greater levels of application within the scope of this thesis. Flow and interaction are critical for work oriented applications. Intuitive and easily understood interfaces that guide the user's experience can increase efficiency and garner a positive user engagement with the product. In the already bustling day to day life within an operational military squadron, with high demands on personnel's time, efficient use of time is critical. No instructor or member of the leadership is interested in a tool that becomes a requirement to use and with which is cumbersome and difficult to interact.

G. CHAPTER II SUMMARY

In this chapter, we first discussed the naval aviation training progression to frame the context of how the naval aviation training is conducted. We then covered the systems approach to training, which is utilized to design and training regimens. The Marine Corps has adopted this and built its T&R program around the concepts are held within the systems approach to training. Then, some underlying theory of instructional design was reviewed to understand the design of instructional systems and their implementation. We then discussed evaluation in greater detail, focusing on the importance of a coherent and relevant evaluation strategy. Next, in order to understand how to be informed by evaluations, what decision support systems are and how they differ from management information systems was covered. This particular research is focused on guiding the development of a decision support system that can be used at the squadron-level to aid in decisions regarding aviator training. We also discussed multiple decision support system development models that provide a foundation with which to classify information that is required in order to ensure the full development of the tool and provide a means to consider how the tool will be used. Mathematical models were reviewed that could form the basis for a decision support tool for Marine Aviation. The most promising mathematical models were the CaRBS and QCA models. Finally, the theory and design

of user interfaces was discussed providing a foundation for a product that is intuitive, desirable by the user, and displays relevant information that can provide insights without further manipulation.

III. METHODOLOGY

This chapter discusses the collection of relevant data from active duty squadrons stationed aboard Marine Corps Air Station Camp Pendleton, California, including the development of a survey that polled IPs within MAG-39 on their perceptions and recommendations regarding the ATF and the current evaluation system. The usage of the ATF and a detailed description of the meaning behind each standard graded item on the ATF are also provided. After collection the raw data was filtered and analyzed to make inferences about which metrics are most critical as well as which metrics should be incorporated into an informational tool that could be developed to inform trainers and provide ability to provide training intervention when necessary. This information could be used to enhance leadership's understanding of the level of training being conducted and the resulting readiness.

A. COLLECTION OF SAMPLE ATF DATA

In order to address the research questions presented in this thesis, performance data was collected from two operational Marine light attack helicopter squadrons stationed at Marine Corps Air Station, Camp Pendleton, California. Approval was provided by squadron commanders to access the full ATF records of all the pilots assigned to their squadrons. Subsequent approval was obtained by the Naval Postgraduate School Institutional Review Board to conduct the collection of information that contained some minimal personally identifiable information.

Over the course of five days all available individual pilot ATF records of both attack helicopter pilots flying the AH-1W Super Cobra or the AH-1Z Viper and utility helicopter pilots flying the UH-1Y Venom were scanned and saved as PDF files, and encrypted to be transported and analyzed at a later time. These records ranged in length from approximately 100 to 300 pages consisting of all of the completed ATFs for each individual aviator. At the completion, a total of 113 records were collected from the two squadrons. See Figure 5 for pilot type breakdown by percentage.

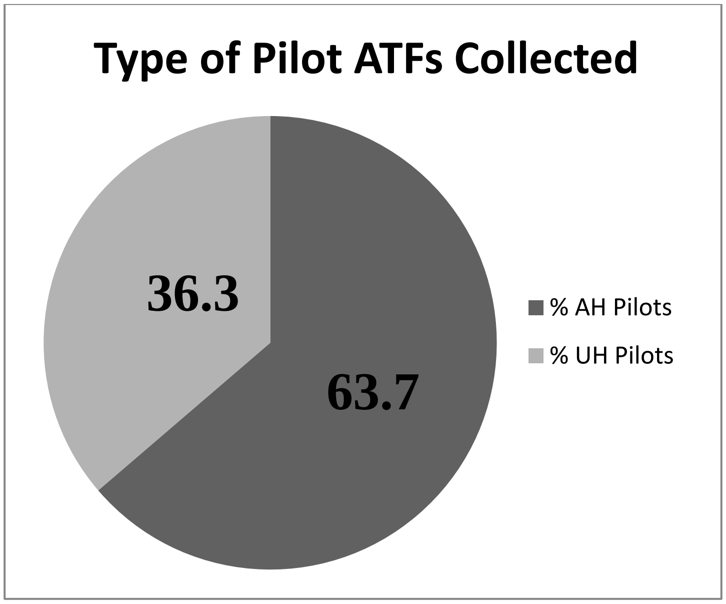


Figure 5. Percentage of pilot type for ATF records collected

The ATF records provided a wide-range completion of syllabus events, since the sample population contained both more senior aviators who had completed most syllabus events, and in some cases completed events more than once, because some had left operational flying and returned, and junior aviators who had only completed one or two events in the Core Skill Phase of training in their current squadron.

1. Processing ATF Data for Analysis

Several challenges existed once the ATFs were saved as PDF files. In order to access and manipulate the data contained within the files, they required conversion to a file type that could be utilized to analyze numerical data. Attempts were made to convert the PDF files to Microsoft Excel files, plain text files, as well as Microsoft document files. None of these attempts were successful due to the variation in which the electronic forms were initially filled out by IPs as well as the variation in which they were printed for retention in the PUIs APR (some were printed in multiple page landscape and others were not). The variation in type setting and the usage of non-standard characters were also used when entering marks on the ATFs making the use of optical character recognition software to process the data inefficient if not impossible. This resulted in the requirement to manually transfer the data into Microsoft Excel for analysis. Further

complicating the analysis of the graded data was the inconsistent format of ATFs throughout the sample of records. Prior to 2011, Marine aviation used an ATF that utilized a different grading scale. The previous scale was a normative scale that allowed for instructors to subjectively evaluate the performance of the PUI by ranking each item as unsatisfactory, below average, average, or above average, no numerical grade was calculated or assigned. Successful completion of the syllabus event was and still remains up to the discretion of the IP under both formats. It should be noted that despite a numerical criterion-based scale, no minimum grade is required to progress. Progression is solely based on the discretion of the IP. The records utilizing the outdated format of ATF were not utilized in any of the analysis conducted for this research, as this would have required these events to be manually and subjectively converted to the new grading scheme.

The remaining records were then screened for completeness and the decision was made to take a sample of ATFs from each aviator's record across his or her performance within the FRS and their operational squadron. Transcribing all ATFs was prohibitively time- and manpower-intensive. The grading data from each of the records selected was manually transferred to an Excel spreadsheet for analysis. This resulted in a sample population of 28 AH pilots and 21 UH pilots and is graphically depicted in Figure 6.

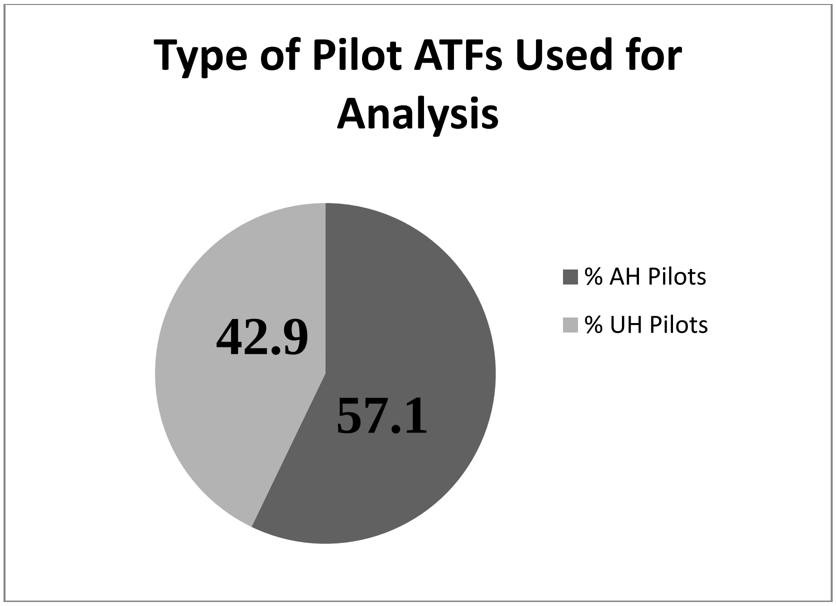


Figure 6. Percentage of pilot type for ATF records used in analysis

For both AH and UH pilots, five flights were taken from their respective FRS syllabi, and 10 flights were taken from their respective Core Skill and Mission Skill syllabus phases. The flights analyzed included only flights up to the point in the syllabus where PUIs were considered competent aircraft commanders who possessed the skill and knowledge to tactically employ their respective platform. The FRS events chosen reflect the middle stages of learning how to maneuver the aircraft, understanding its systems, and how to operate weapons systems. The Core Skill and Mission Skill flights included in the sample data include the PUI's first flight in the squadron and a representative group of flights reflecting both the progression of the PUI and representative tasks that pilots are expected to perform at satisfactory levels. A more detailed explanation of each event selected for analysis can be seen in Table 3.

FLIGHT PHASE	AH EVENT & DESCRIPTION	UH EVENT & DESCRIPTION
Core Skill Introduction (FRS)	FAM-1110: Familiarization Flight consisting of basic aircraft maneuvers and emergency procedures.	FAM-1110: Familiarization Flight consisting of basic aircraft maneuvers and emergency procedures.
Core Skill Introduction (FRS)	FAM-1117: Introductory NVD flight conducting basic aircraft maneuvers.	FAM-1114: Introductory NVD flight conducting basic aircraft maneuvers.
Core Skill Introduction (FRS)	FORM-1303: Introductory NVD formation flight.	FORM-1303: Introductory NVD formation flight.
Core Skill Introduction (FRS)	SWD-1602: Introduce basic conventional weappons delivery (rocket and gun delivery).	SWD-1603: Introduce basic conventional weappons delivery (rocket and gun delivery).
Core Skill Introduction (FRS)	SWD-1605: Weapon system evaluation. PUI shall have detailed understanding and functional knowledge of weapons procedures and checklists.	ASPT-1802: Introduction to confined area landings (CALs), and assault support techniques.
Core Skill	TERF-2100: First flight in squadron. Review terrain flight maneuvers and conduct a navigation route.	TERF-2100: First flight in squadron. Review terrain flight maneuvers and conduct a navigation route.
Core Skill	REC-2300: Introduction to daytime visual reconniassance.	ASPT-2400: Introduction to section tactical landings and tactical approaches.
Core Skill	SWD-2602: Specific weapons delivery and employment of hellfire missile system with a live missile.	SWD-2603: Proficiency building for specific ordnance delivery (rockets and guns).
Core Skill	SWD-2604: Proficiency building for specific ordnance delivery (rockets and guns).	SWD-2605: Proficiency evaluation for specific ordnance delivery (rockets and guns).
Core Skill	SWD-2607: Refinement of ordnance delivery using NVDs under high light level (HLL) conditions (rockets and guns)	SWD-2607: Refinement and proficiency building of ordnance delivery using NVDs under high light level (HLL) conditions (rockets and guns)
Core Skill	ANSQ-2705: Review ordnance delivery under low light level (LLL) conditions.	ANSQ-2703: Review of navigation, tactical landings and ordnance delivery under LLL conditions.
Mission Skill	ESC-3103: Introduction to surface force escort in a low to medium threat environment.	ESC-3103: Introduction to surface force escort in a low to medium threat environment.
Mission Skill	CAS-3303: Provide close air support (CAS) to ground forces in a medium threat environment.	AD-3205: Tactical employment of aircraft in support of a raid, insert or extract mission with a follow on resupply.
Mission Skill	AI-3306: Conduct an air interdiction (AI) mission in a medium threat environment.	CAS-3303: Provide close air support (CAS) to ground forces in a medium threat environment.
Mission Skill Designation	AHC-6398: Evaluation flight resulting in designation as an aircraft commander. PUI demonstrates all required skills of Core Skill and Mission Skill phases.	UHC-6398: Evaluation flight resulting in designation as an aircraft commander. PUI demonstrates all required skills of Core Skill and Mission Skill phases.

Table 3. Syllabus events utilized for analysis for AH and UH aircraft from respective training and readiness manuals.

The events outlined in Table 3 are intended to capture a representative collection of training events conducted throughout the progression of a PUI through their respective syllabi. These events gradually build in complexity and increased responsibility for the PUI. Aviators are expected to continue to progress through further events that focus on more advanced mission skill sets and flight leadership events after their designation as an aircraft commander. The Core Skill and Mission Skill phases provide aviators with the foundational knowledge and experience to progress to these more advanced events. The

assumption by the author is that this foundational experience should be sufficient to examine trends and identify evaluated items that may be most influential in performance prediction.

B. CREATION OF INSTRUCTOR PILOT OPINION SURVEY

In order to better inform the research, a survey was devised to collect data on the opinions of those aviators tasked with instructing and evaluating PUIs, and filling out ATFs to communicate the status of the individuals they trained. Permission was obtained from a Marine Air Group (MAG) to electronically survey all helicopter pilots who possessed an instructor qualification. Participation in the survey was voluntary. Solicitation for participation was conducted via email. The survey was available through a LimeSurvey internet site for a period of four weeks with a re-solicitation after two weeks to provide a reminder to potential participants who had not yet completed the survey. Ideally, a survey of all Marine aviators possessing an instructor qualification would have been conducted. It is believed that the opinions collected across a single MAG span a representative range of instructor experience and opinions that will present themselves in other MAGs across the Marine Corps regardless of type of aircraft flown or location of the particular unit.

The survey (see Appendix B), created using the LimeSurvey tool available to Naval Postgraduate School researchers, collected demographic information about participants including total hours flown and which instructor qualifications they possessed, as well as their opinions regarding the training and readiness manual for their respective type, model, and series of aircraft and ATFs. The LimeSurvey online tool allowed for automated data collection and reduced the time required for travel to conduct surveys as well as to transfer data from paper copies to electronic format. Collecting information on which qualifications are held by each participant allowed the investigation of how the importance of items changed across the levels of instructor experience. The survey "Instructor Pilot Attitudes Toward Current ATF Ratings" asked a series of 12 questions soliciting the instructors' opinions on the importance of the standard items, mission-specific items and remarks and comments provided by the current form of the

ATF in use. The survey also provided a free response section to allow participants to make recommendations on what features they were interested in having available in a tool developed to aid in the evaluation and assessment of PUIs (see Figure 7).

Figure 7. Example of free response question

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA ANALYSIS AND TOOL DESIGN

This chapter outlines the data analysis conducted on the ATF and survey data collected. The analysis was conducted to support the development of an instructional tool that informs training decisions at the squadron level. Analysis is required in order to make inferences regarding personnel performance and provide a means with which to make sense of performance data for decision makers. This information also has the potential to further inform upper levels of command on the qualitative level of instruction being conducted at the subordinate units.

A. ANALYSIS OF AVIATION TRAINING FORM PERFORMANCE DATA

The first analysis conducted of aviation training data examines the descriptive statistics of the aggregate performance found in the sample data. For the purposes of this research, we have assumed that data collected is sufficiently representative of the total population of both AH and UH pilots that have been trained. Figure 8 provides a look at the distribution of average AH and UH pilot grades across the 10 standard ATF grading metrics. Summary statistics for this distribution are listed in Table 4.

Distribution of AH and UH Overall Averages

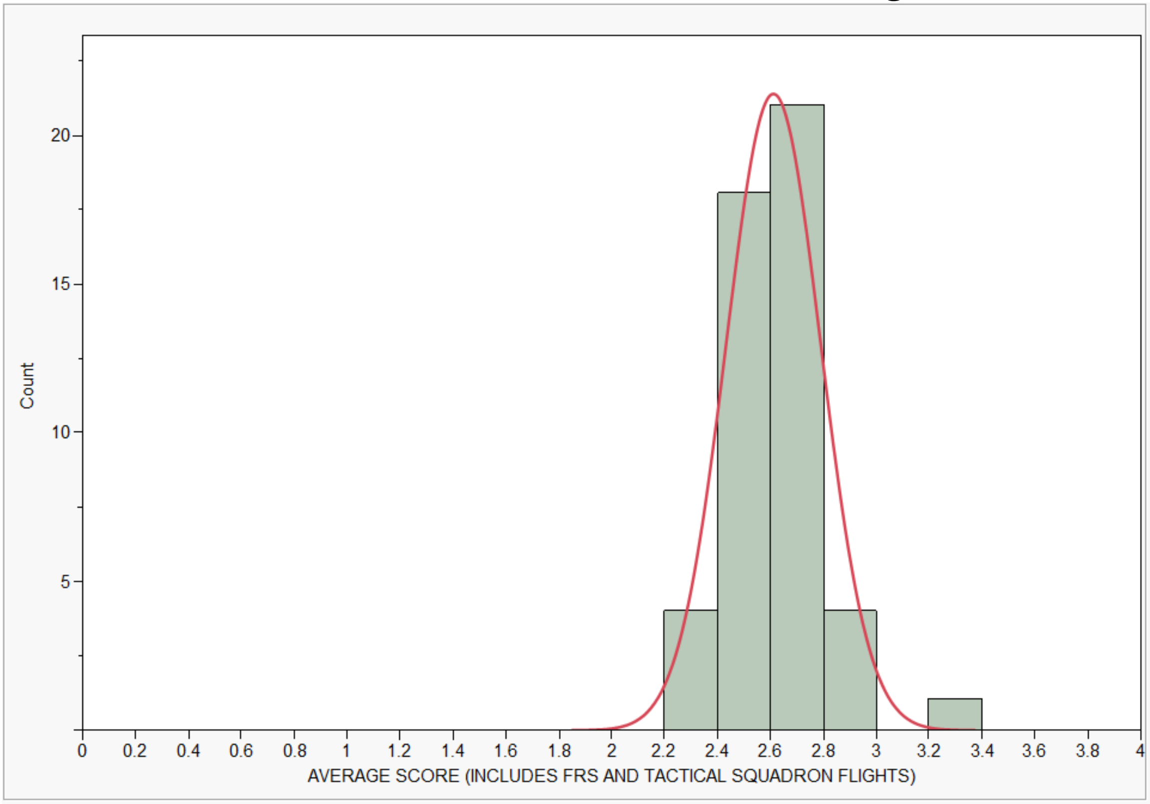


Figure 8. Distribution of AH and UH overall grades among sample ATFs with a fitted normal curve overlay

Overall Average Summary Statistics (Confidence Level = 0.95)	
Mean:	2.61
Std Deviation:	0.18
Lower Confidence Level:	2.56
Upper Confidence Level:	2.66

Table 4. Summary statistics for the distribution of overall event averages (n = 48)

Assuming a normal distribution, and given the sample, with 95 percent confidence we could expect the true mean of aviator grades on the standard ATF items to fall between 2.55 and 2.66. This provides an overall baseline with which to compare individual performance to the population. The mean scores specific to FRS training and tactical squadron training were also examined separately (see Table 5 for summary statistics). The scores achieved by the sample population in the FRS have a mean of 2.99,

and with 95 percent confidence we expect the true mean of aviator grades in the FRS on the standard ATF items to fall between 2.93 and 3.06 (see Figure 9).

Distribution of AH and UH FRS Averages

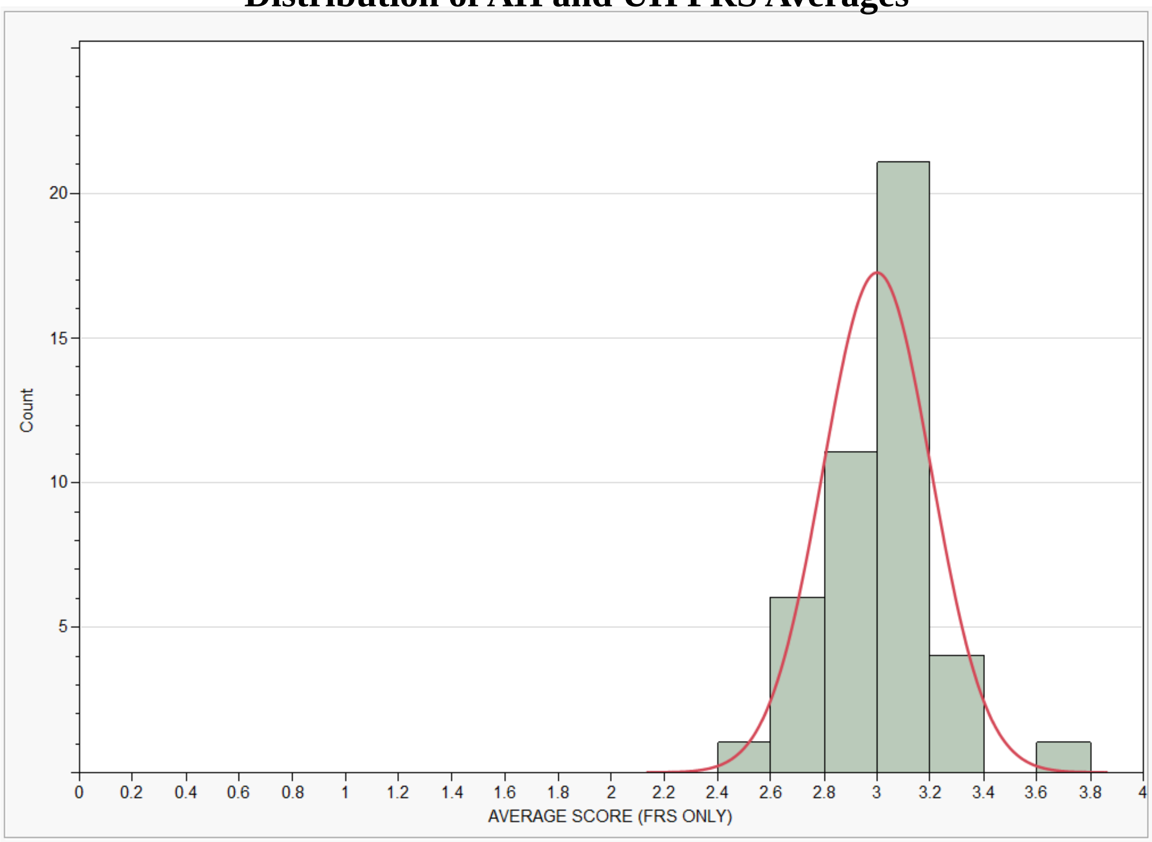


Figure 9. Distribution of AH and UH FRS-only grades among sample ATFs with a fitted normal curve overlay

FRS Average Summary Statistics (Confidence Level = 0.95)	
Mean:	3.00
Std Deviation:	0.20
Lower Confidence Level:	2.94
Upper Confidence Level:	3.06

Table 5. Summary statistics for the distribution of FRS averages (n = 44).

The scores achieved by the sample population in the squadron have a mean of 2.39, and with 95 percent confidence we expect the true mean of aviator grades in the squadron on the standard ATF items to fall between 2.33 and 2.44 (see Figure 10 and Table 6).

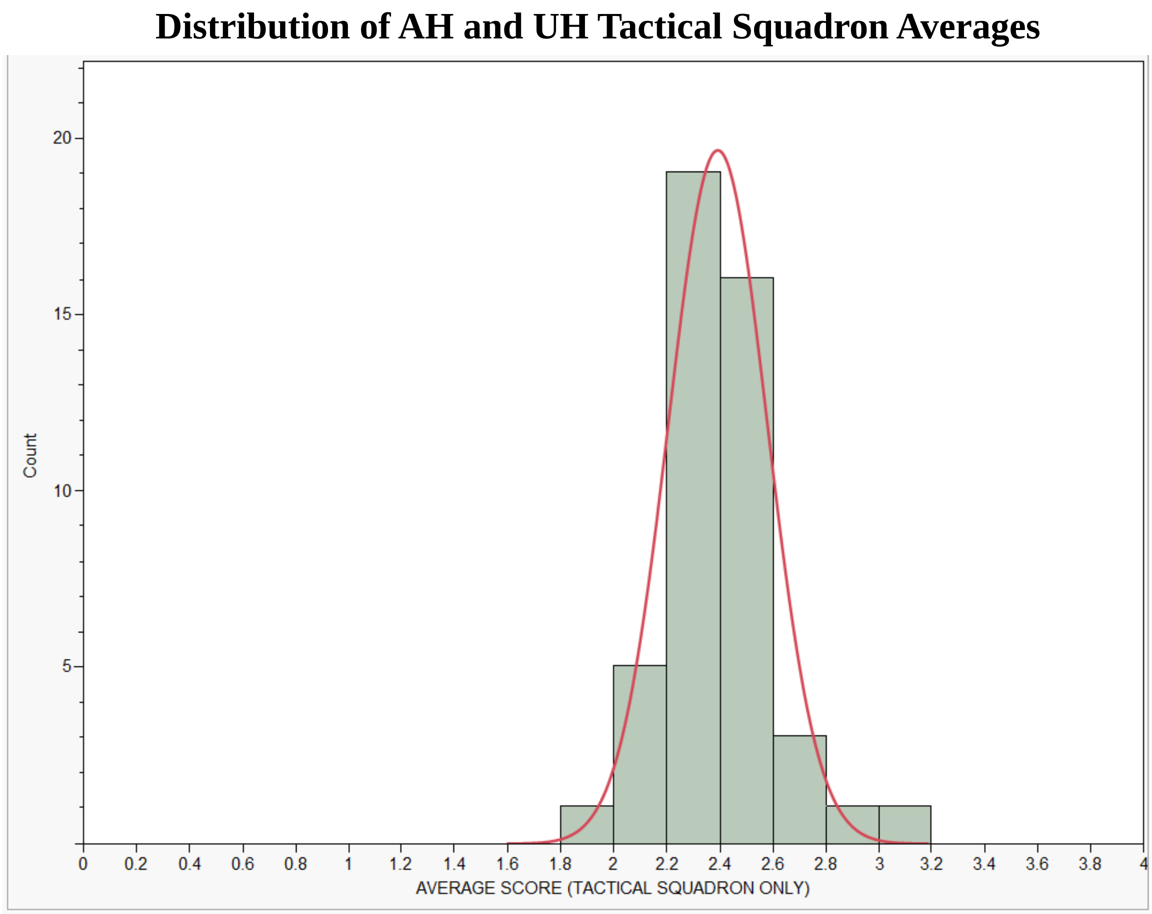


Figure 10. Distribution of AH and UH Squadron-only grades among sample ATFs with a fitted normal curve overlay

Tactical Squadron Average Summary Statistics (Confidence Level = 0.95)	
Mean:	2.39
Std Deviation:	0.19
Lower Confidence Level:	2.34
Upper Confidence Level:	2.45

Table 6. Summary statistics for the distribution of tactical squadron averages (n = 46)

The higher mean score achieved at the FRS versus the squadron might be attributable to the difference in criteria for grading found on the FRS ATFs versus the ATFs for training to be conducted at the squadron as discussed in Section C.3. To confirm the difference between the FRS and squadron means a one-way analysis of averages by squadron with which the training was conducted using JMP software. The results can be seen in Figure 11, which confirms via the student's t-test that the population averages over FRS events and events completed in the tactical squadron are different (also see Table 4).

Oneway Analysis of FRS and Tactical Squadron Averages

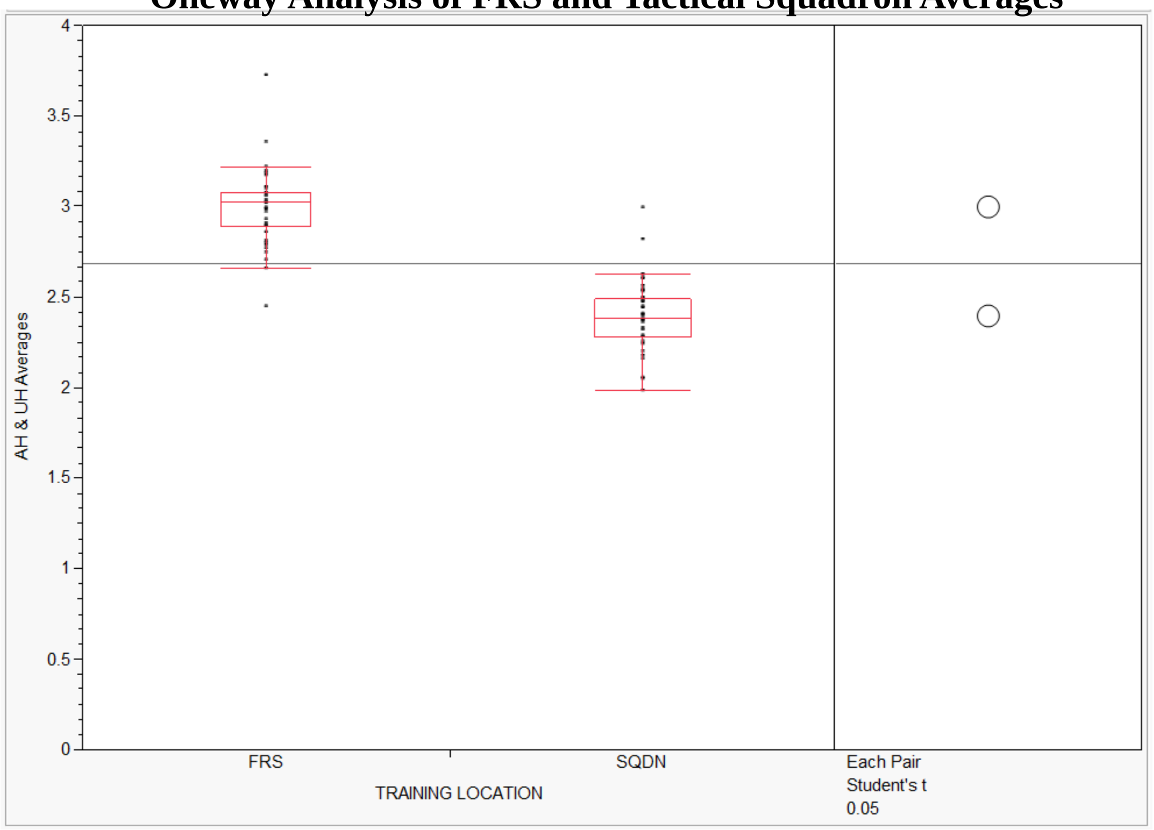


Figure 11. Means comparison of FRS and squadron ATF averages.

Detailed Means Comparison Report for Averages by Squadron	
FRS Mean:	3.00
FRS Std Deviation:	0.203
Tactical Squadron Mean:	2.39
Tactical Squadron Std Deviation:	0.187
Difference of Means (FRS-Tactical Squadron):	0.610
t-ratio:	-2.37
p-Value:	< 0.0001

Table 7. Detailed Means Comparison Report for Averages by Squadron Type (FRS n = 44, Tactical Squadron n = 46)

The statistically significant difference between the FRS and the squadron is important to understand if we are to use these values as a baseline with which to compare the performance in a population. It may be useful to separate these averages when using them as a baseline, in order to minimize the interactions between the slightly dissimilar criterion references found on the Core Skill Introduction Phase conducted at the FRS and the subsequent phases conducted at the tactical squadron.

The means were also compared by pilot type. The null hypothesis in this case is that both AH and UH pilots have the same average scores over the course of equivalent training stages. The comparison is displayed in Figure 12. Utilizing the Student's-t each pair comparison, a p-value of 0.022 was computed (see Table 5), and we conclude that the population averages are significantly different in statistical terms. This is an interesting result given that the PUIs are executing equivalent events. The sample provides data that suggests that AH pilot averages are higher than UH pilots. Because of several confounding influence factors we cannot assert the reason. Some possible reasons might include that AH IPs are more lenient, UH IPs are less lenient, AH PUIs are slightly more capable than UH PUIs at equivalent stages, the events are less difficult for AH PUIs and more difficult for UH PUIs resulting in higher grades for the former, or even that the differences in cockpit configuration (tandem in AH-1 cockpits, and abreast in UH-1 cockpits) result in grading differences.

Oneway Analysis of AH and UH Overall Averages by Pilot Type

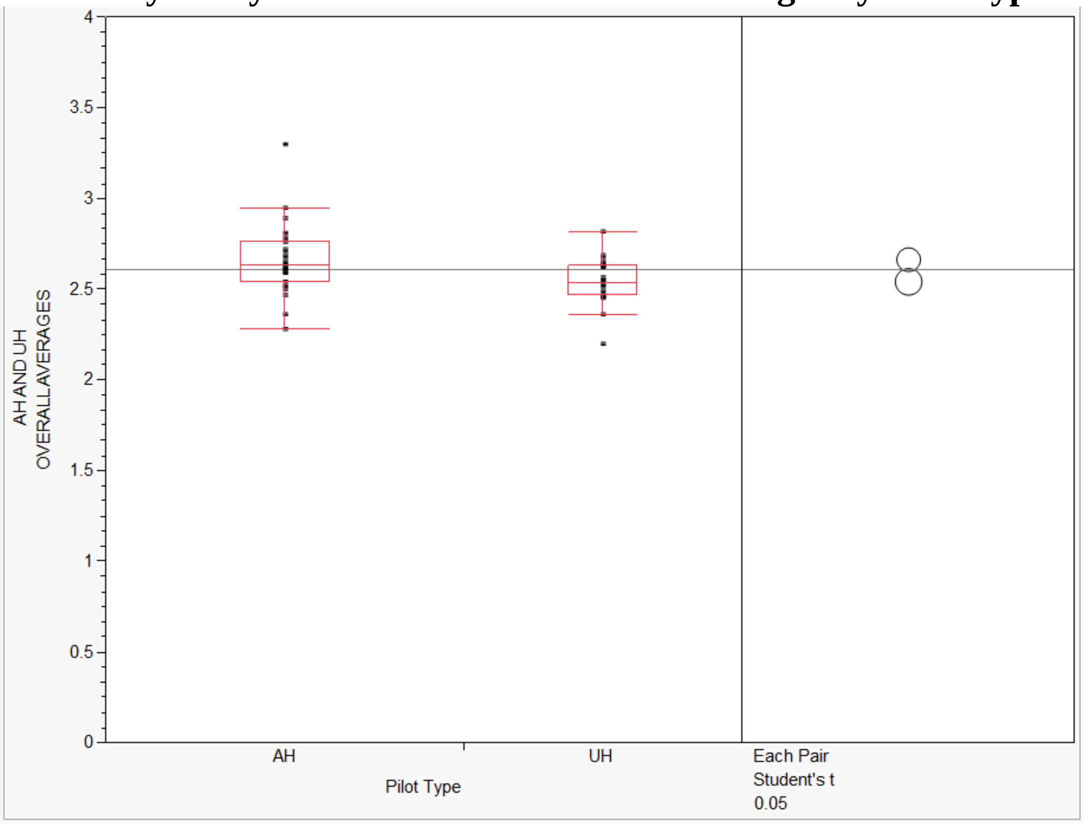


Figure 12. Means comparison of AH and UH pilot ATF averages.

Detailed Means Comparison Report for Averages by Pilot Type	
AH Mean:	2.66
AH Std Deviation:	0.196
UH Mean:	2.54
UH Std Deviation	0.131
Difference of Means (AH-UH):	0.118
t-ratio:	-2.37
p-Value:	0.022

Table 8. Detailed Means Comparison Report for Averages by Pilot Type (AH n = 27, UH n = 21)

The next step in the analysis was to examine the individual metrics and their effect on the overall averages that were achieved. First, the average of each graded item was computed for the entire sample, as well as across the sample of AH and UH pilots. The averages are plotted in Figure 13.

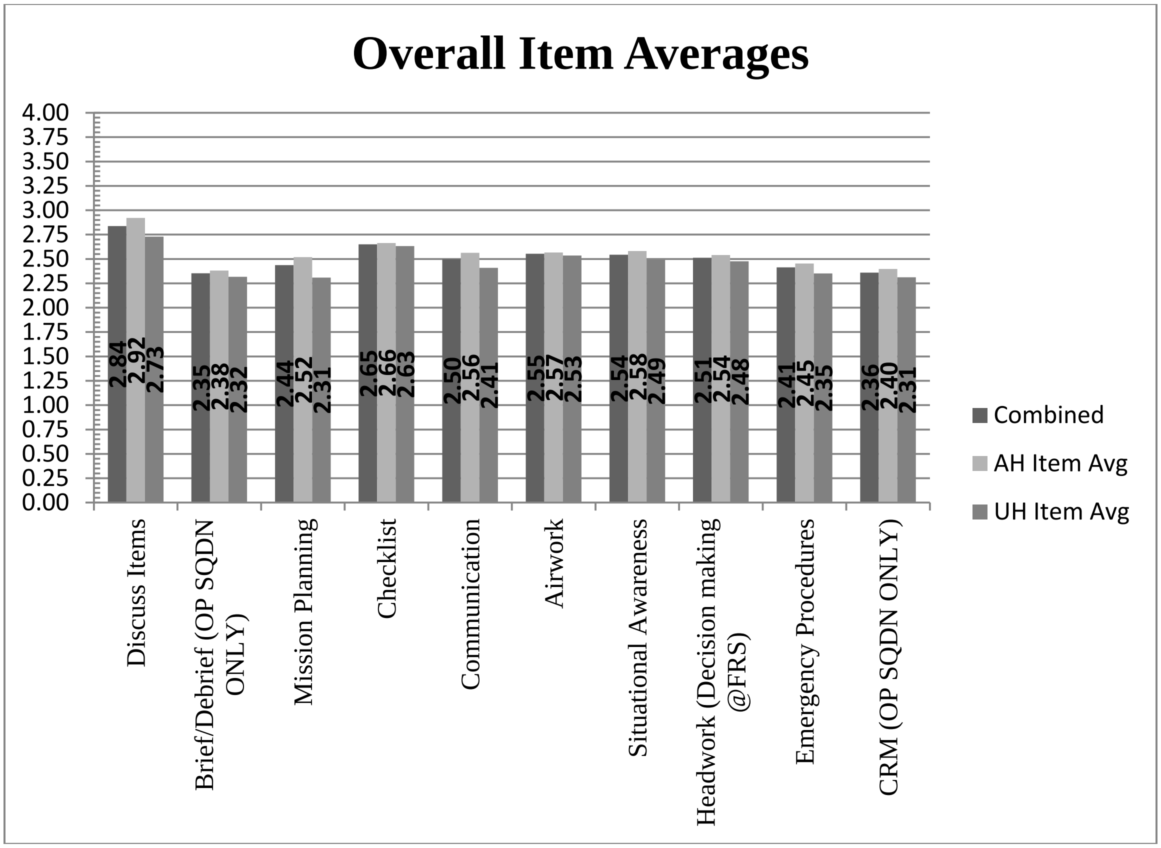


Figure 13. Plot of averages by specific grading item

When the combined averages are plotted and examined in order from highest to lowest, we see that the items with the highest averages are discuss items and checklist (see Figure 14). This result makes sense because discussion items and checklists are the most basic tasks that a PUI is expected to perform. The discussion items are delineated in the T&R manual and generally discussed between the PUI and the IP for an upcoming event prior to the scheduled activity. An analogy would be a teacher telling a class what topics to review in a textbook prior to a quiz or test. Checklists are drilled continuously from the earliest stages in flight training and, after initial exposure to the aircraft checklist in the FRS, PUIs are expected to be able to efficiently and effectively follow the checklist that is read from, item by item as required, as the aircraft is started, flown, landed and shut down.

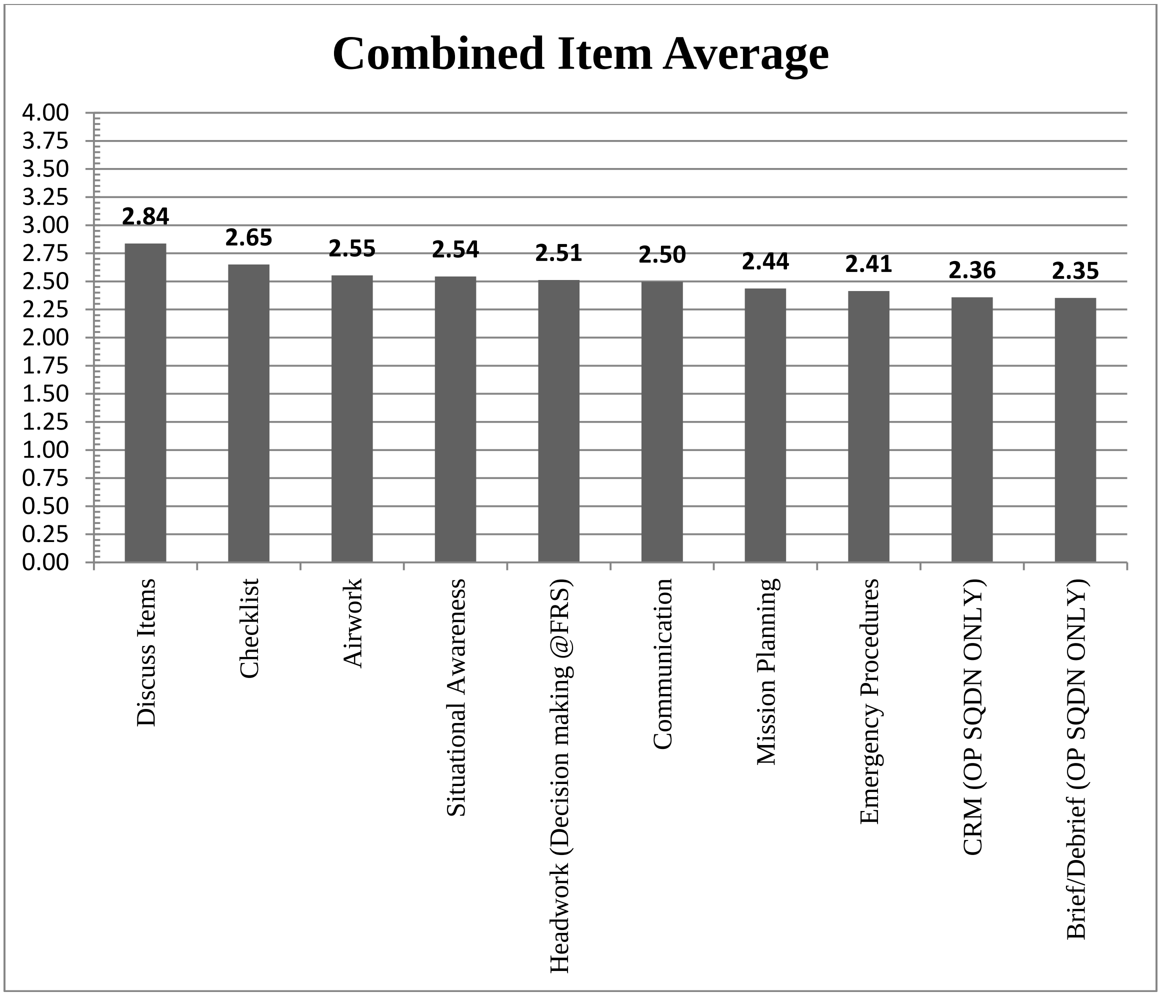


Figure 14. Ordered plot of combined item averages

The item on which PUIs had the lowest score, Brief/Debrief, is an item on which PUIs are heavily scrutinized in a setting that is generally most conducive to note-taking by IPs, and is less time-sensitive than items that are graded based on in-aircraft performance. The item with the second lowest average, CRM, is graded at the operational squadron only (it is broken into its component parts and each component is marked individually at the FRS, see Table 2). This could be due to a number of different factors, which include a PUI's inability to mentally keep up in a new tactical environment to which pilots will not have been exposed in their aviation careers, and a lack of understanding of how to meaningfully participate as a crewmember in the tactical environment.

Another phenomenon that was recognized across the ATF data was that particular instructors appear to have a typical grading profile for a specific event. For example the instructor identified as C4, who instructed the simulated specific weapons delivery event while PUIs were at the FRS, assigned the exact same grades five of six times. In each instance that was identical, the IP only assigned grades for discussion items, checklists, airwork, communications, and situational awareness. In the dissimilar score set the instructor changed the score received by the PUI for airwork by one level decrease, did not assign a score for communication, assigned scores for mission analysis, adaptability and flexibility, and emergency procedures. This can also be seen in the sample data from instructor M129 on an escort mission event where the same grades were assigned for all five events flown by a different PUI. This raises the question as to whether these grades are meaningful if all PUIs receive the same score. The particular event referenced here was conducted in the simulator. If all events were able to be analyzed in greater detail this may be true for more events. The question then becomes whether grades for events graded in such a manner even require grades to be assigned, and whether they should be considered on a pass or fail basis only. This is especially true if these grades are being used by decision makers on how PUIs are progressing.

Finally, we examined the occurrence of grades at the extremes of the ATF grading categorical scale. The percentage of grades assigned at the upper level of the scale was 4.4 percent with only 2.7 percent of scores assigned at the low end of the scale, but not considered unsatisfactory. Figure 15 provides a graphical representation of these two extremes and their occurrences by ATF item. It clearly shows that discussion items received the largest number of highest marks, while the low scores were more evenly distributed between airwork, situational awareness, discussion items, checklists, and communication (also see Table 4).

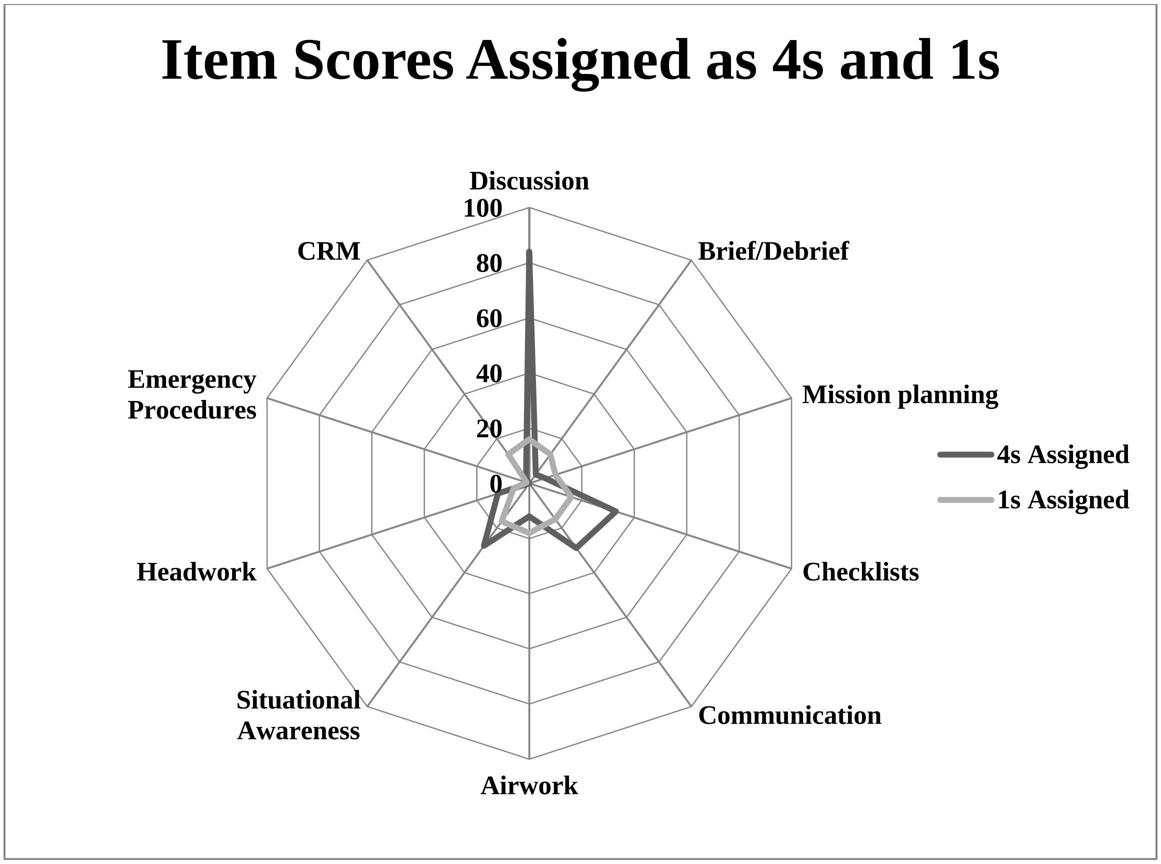


Figure 15. Radar plot of item scores assigned at the extremes of the ATF criterion scale

	4s Assigned	**1s Assigned**	**Total Graded Items**	**Percentage 4s Assigned**	**Percentage 1s Assigned**
Discussion	84	16	589	14.3%	2.7%
Brief/Debrief	4	13	337	1.2%	3.9%
Mission planning	6	10	417	1.4%	2.4%
Checklists	33	16	569	5.8%	2.8%
Communication	29	16	596	4.9%	2.7%
Airwork	12	18	584	2.1%	3.1%
Situational Awareness	28	17	591	4.7%	2.9%
Headwork	12	6	543	2.2%	1.1%
Emergency Procedures	0	1	140	0.0%	0.7%
CRM	2	13	384	0.5%	3.4%
Totals	210	126	4750	4.4%	2.7%

Table 9. Table of scores assigned at the extremes of the ATF criterion referenced scale

This information allows us to make a number of possible inferences. One such inference is that most instructors generally assign a grade, independent of item, at the center of the scale. Another item of note is that despite the assignment of one flight from the dataset being graded as unsatisfactory, no ATF items on any flights were graded as such. It also raises the question as to whether most PUIs perform at the center of the scale. We also might be able to explain the high number of high marks on discussion items by understanding that this item is essentially a recitation of items expected to be studied by the PUI. The results also might indicate that there are more PUIs that exhibit a high degree of ability and require no further instruction on that item than those that have limited proficiency and require frequent instructor input.

B. ANALYSIS OF INSTRUCTOR PILOT SURVEY RESPONSES

The survey data collected was collected from voluntary participants who held instructor designations in MAG-39. The sample population included instructor pilots of transport, utility, and cargo-carrying helicopters. A recruitment email (see Appendix C)

was distributed via the global address list on the USMC dot-mil enterprise email network to squadron instructor distribution lists. After the initial email a second email was sent after a two-week period to remind potential participants that the survey was still accessible and could be filled out. At the completion of the survey period, 34 participants submitted complete responses. Incomplete responses might be attributed to respondent unwillingness to fill out written portions of the survey, or a change in decision to participate mid-survey.

1. Demographic Information

The first two questions of the survey focused on demographic information. The first question asked the participant to indicate all of the instructor qualifications that they held. This question revealed that of the IPs who participated, a majority of them held a senior-level qualification, namely night systems instructor (NSI). Table 10 displays qualifications held by participants by count and by percentage of total responses.

Qualification	Count (out of 34 total Responses)	Percentage of Total Responses
BIP	31	91.2%
TERFI	32	94.1%
WTO	30	88.2%
TSI	21	61.8%
NSI	24	70.6%
DACMI	8	23.5%
FAC(A)I	9	26.5%
FLSE	11	32.4%
FRSI	8	23.5%
NSFI	7	20.6%

Table 10. Survey qualification demographics

The qualifications listed from NSI and below in Table 10 indicate that these instructors are capable of training inexperienced PUIs under the challenging conditions of nighttime flying while employing weapon systems. Holding this qualification also implies that the IPs' command hold considerable trust and confidence in them, since obtaining the qualification of NSI requires considerable internal and external training resources.

The second question of the survey asked participants to report their total career military flight hours flown. This value includes those flight hours flown in the primary, intermediate, and advanced stages of naval aviation training. The results are displayed in Figure 16 and Figure 17.

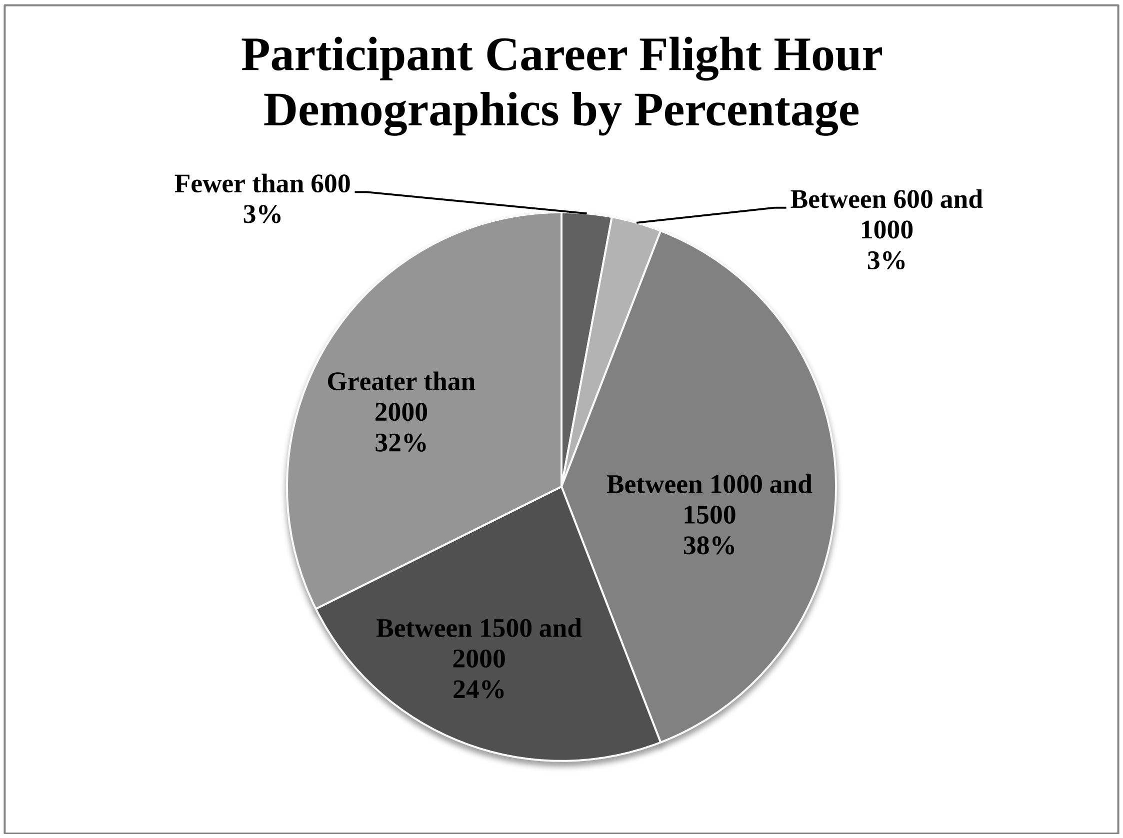


Figure 16. Survey participant career flight hours by percentage

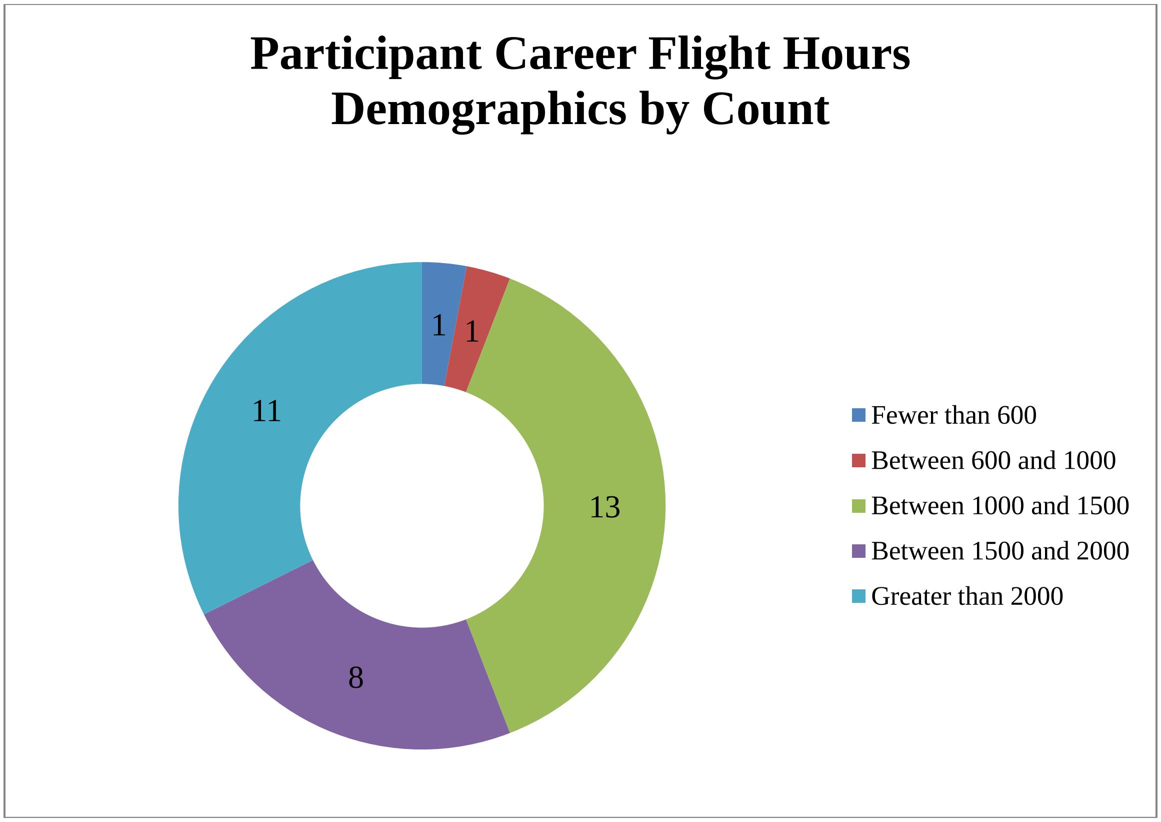


Figure 17. Survey participant career flight hours by count

Of the 34 participants, 55.9 percent have flown over 1500 career military flight hours. Again, this information allows us to infer that the participants, in general, have a considerable amount of collective experience, despite the limited sample size.

2. Pilot Opinion Data

Following the demographic questions, participants were asked to provide responses to questions designed to elicit their opinion on several aspects of the ATF and the system of evaluation in use by Marine aviation. The first question of the survey related to instructor opinion asks whether the participant believes the T&R manual for their respective T/M/S of aircraft clearly defines the standards to which PUIs are expected to perform. 70 Percent of participants consider the performance standards to be clearly defined in the T&R manual (see Figure 18).

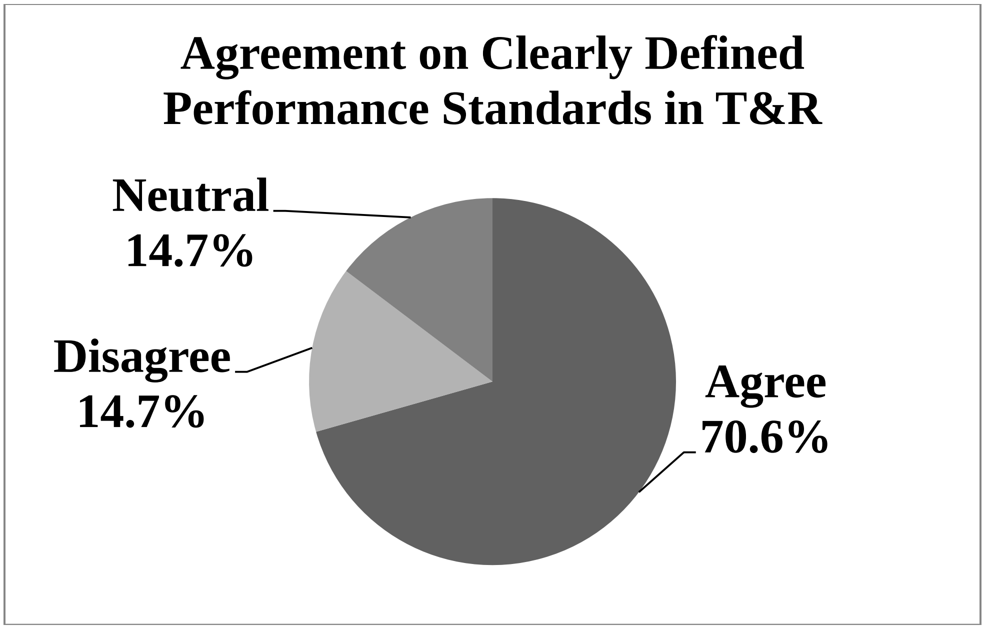


Figure 18. Percentage of participants who agree or disagree with clearly defined performance standards in the T&R manual

This suggests that a majority of instructors who participated in the survey believe that the standards to which PUIs are expected to perform are well understood. It cannot be determined from the survey whether the participants themselves understand the performance standards. Nor can it be determined whether the participants believe that they themselves are proficient at applying these standards when evaluating trainees. Further examination of the responses reveals that of the 71 percent that agree the standards are clearly defined only six percent of those surveyed strongly agree that these standards are clear (see Table 11). It can be inferred from this question that many instructors believe that the performance standards could be more clearly defined within the T&R manual.

Answer	Count	Percentage
Strongly Disagree	1	2.9%
2	1	2.9%
3	3	8.8%
Neutral	5	14.7%
5	13	38.2%
6	9	26.5%
Strongly Agreee	2	5.9%

Table 11. Responses to agreement with statement: "The performance standards in the Training and Readiness Manual for my T/M/S are clearly defined."

The second series of responses asked participants to rate the level of importance for individual graded items of the ATF when assigning scores and when assessing PUIs based on ATF entries (see Figure 19, Figure 20, Figure 21, and Figure 22).

Figure 19. LimeSurvey ATF standard item importance survey question

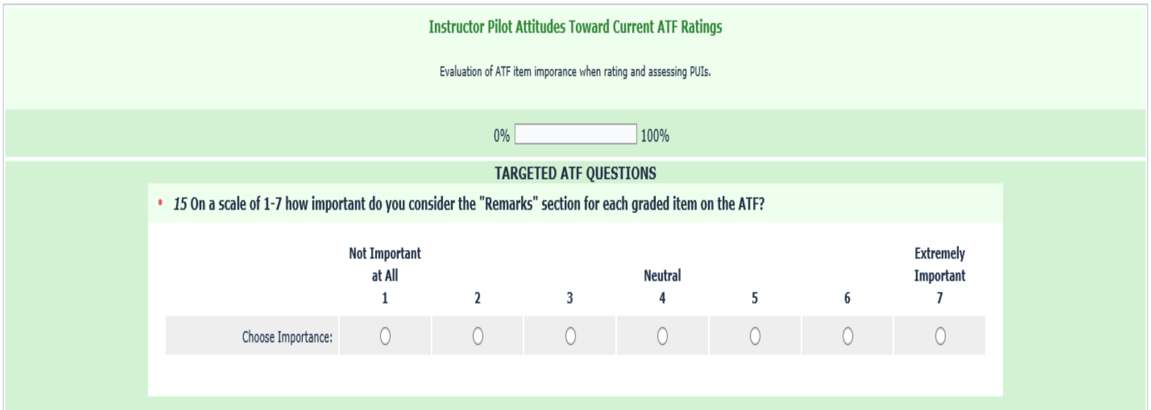


Figure 20. LimeSurvey ATF “Remarks” item importance survey question

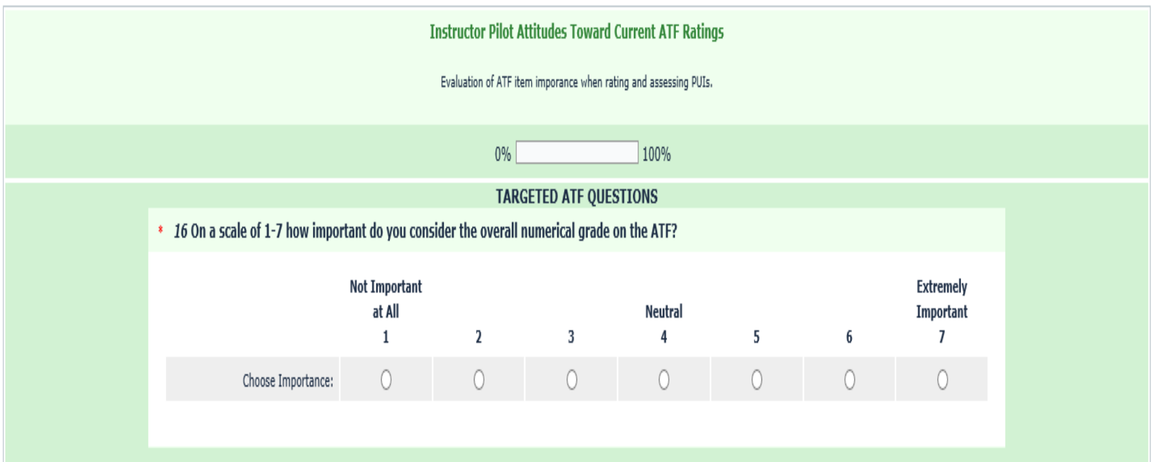


Figure 21. LimeSurvey ATF overall grade importance survey question

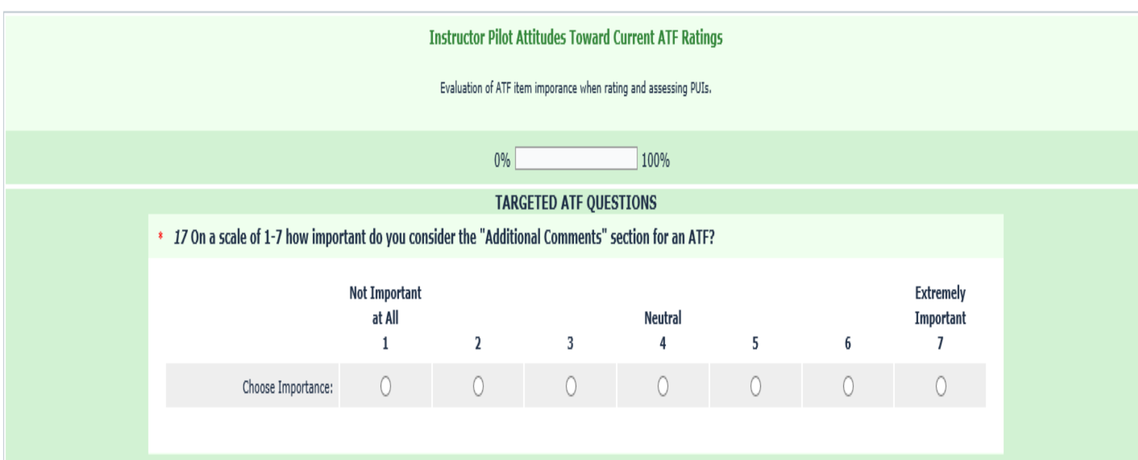


Figure 22. LimeSurvey ATF “Additional Comments” importance survey question

The most important item on the ATF to IPs who participated in the survey when the level of response scores are summed is situational awareness followed by headwork, and the least important item is the overall grade, which is merely an average of all scores obtained on a particular flight. The second least important item is checklists. The ranked order of all items is seen in Figure 23.

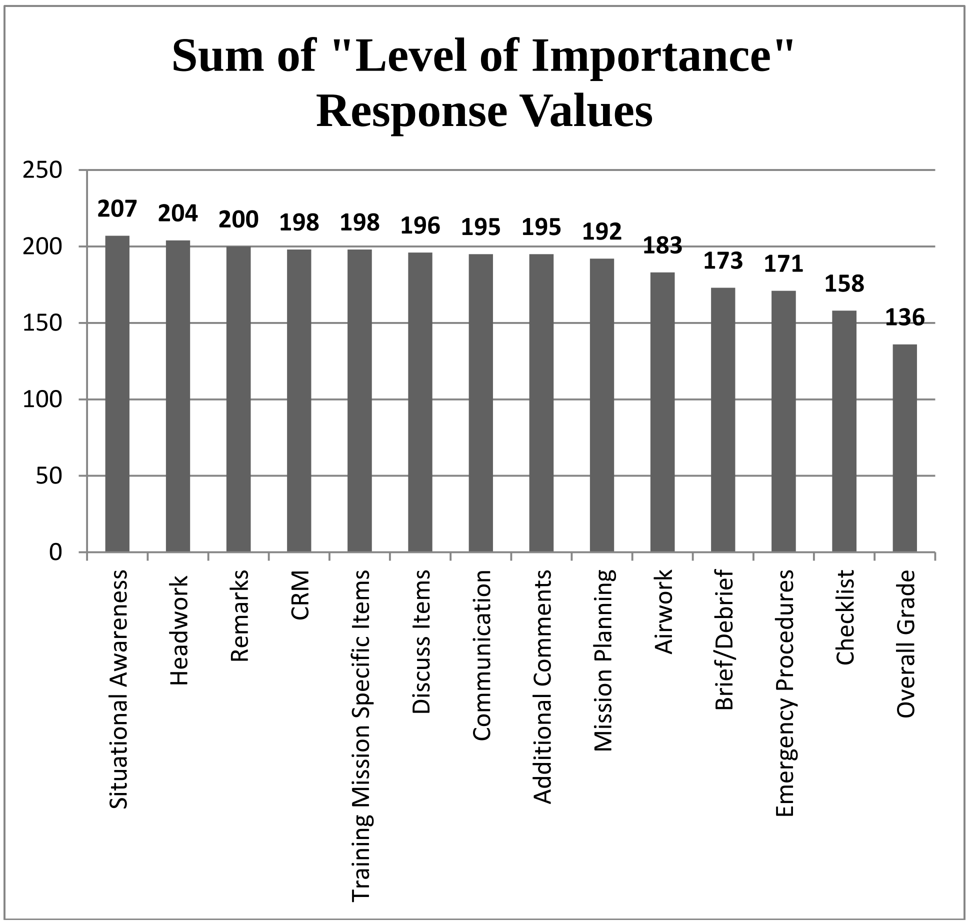


Figure 23. Sum of response values for "Level of Importance" of ATF graded items

Of interest here is that remarks provided by IPs on the ATF is ranked third and CRM is tied for the rank of fourth with training specific items. Situational awareness, which is a principle of CRM, and headwork, defined in Chapter II, Section C.3, are closely aligned in the opinion of IPs when it comes to how the PUI performs on a whole for an individual event. Figure 23 also shows that headwork and additional comments are separated by a spread of 10 points. This suggests that these items are of similar

importance to IPs when determining the performance of a PUI based on the ATF alone. Also of interest is that of the graded items found on the ATF, the three most important items are situational awareness, headwork and CRM. These three items are very difficult to quantify numerically, but are considered the most important to the survey participants. IPs are required to make subjective judgments on how closely a PUI meets the criteria provided on the reference scale provided on the ATF.

The next question asked the participant to indicate his or her level of agreement with a statement regarding flights that evaluate PUIs as requiring additional training (RAT). The question asks specifically for the participant's opinion on whether a grade of RAT is derogatory towards a PUIs performance record or not. The ATF explicitly states that assignment of RAT is not derogatory towards the PUI's record; however, it does indicate that he or she needs more training or exposure in components of the skills being taught on that particular training event. When asked their opinion on whether IPs believe this is true, we find that results are mixed. Figure 24 clearly shows that opinions among respondents are split evenly between those that agree that the current RAT policy holds true when these events are assigned.

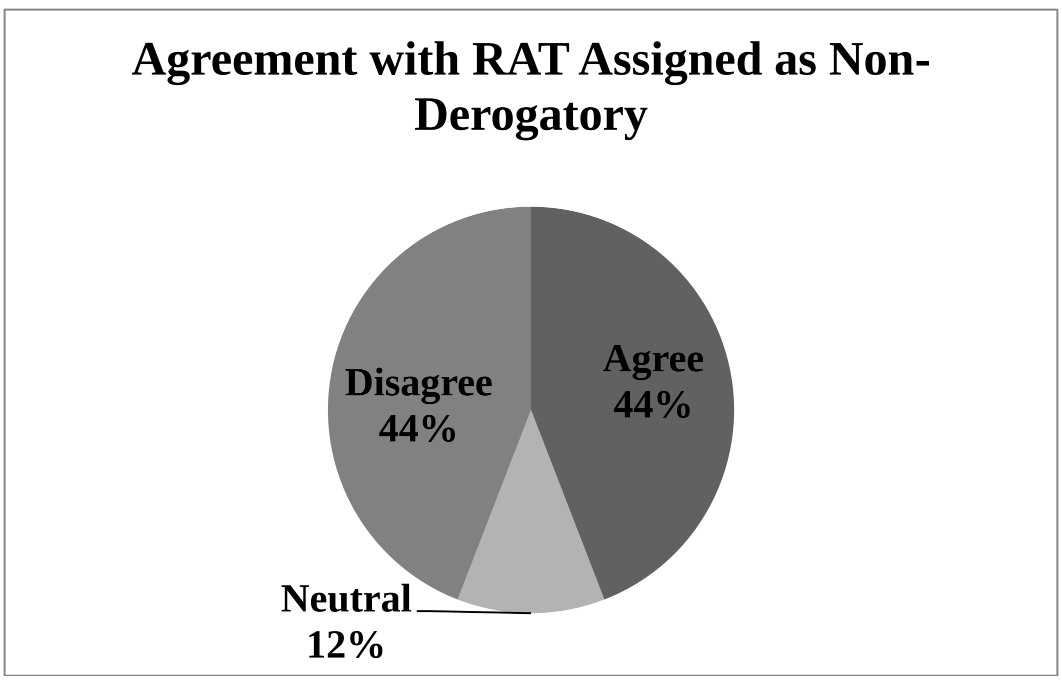


Figure 24. Agreement with RAT assigned as non-derogatory

This split suggests that there is considerable disagreement on whether the assignment of a RAT grade is truly viewed as non-derogatory or if it has some negative impact on the impression left with an IP or leadership when they encounter this grade on an ATF. If nothing more, this result indicates the need for a discussion regarding the merits of a RAT grade and whether it should be an option for IPs. This leads to the question of whether there is a presumption that a PUI can successfully complete a training event before it is assigned on the flight schedule. If this presumption exists, the RAT grade is misplaced, because if PUIs are assigned the RAT they are not keeping up with the expected level of performance. This may be the case since prerequisites for each training flight event are delineated in the T&R. However, if a PUI is assigned to fly the next flight in the syllabus simply because the aircraft, ordnance, and instructor required are available, it is plausible that a PUI could need the aforementioned additional training and exposure.

Question 19 asks the participant to indicate their level of agreement with a statement regarding the completeness of the ATF when it comes to the performance information it provides to evaluators (see Figure 25).

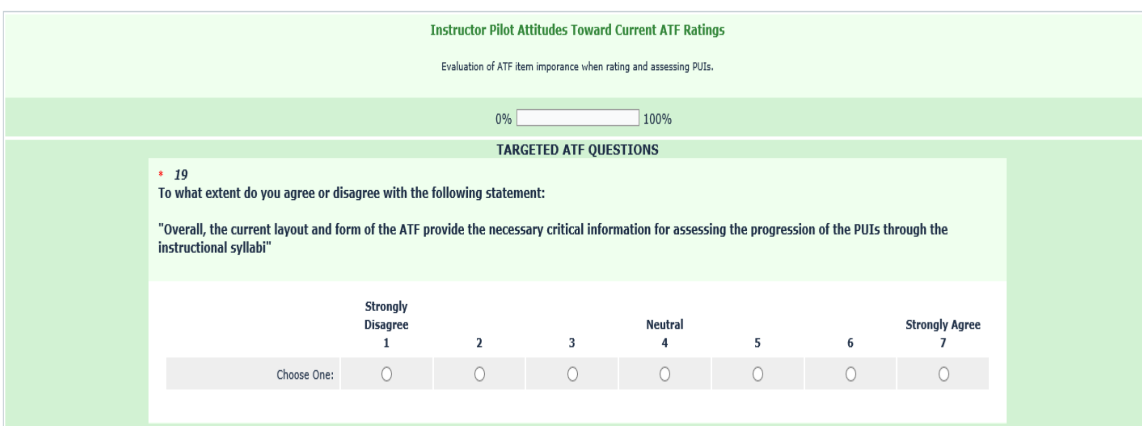


Figure 25. LimeSurvey question regarding completeness of ATF with regards to critical information for evaluation

If a participant's response was at any level of disagreement with the statement posed in the survey item (they selected a number between one and three), they were then asked to enter free text describing what they believed the ATF was lacking. If they provided a neutral response or one in agreement with the statement, they were directed to the next question. Five of the 34 participants disagreed with the statement, and all five responded with a disagreement level of three on the provided scale. Their responses can be viewed in Table 12.

Response ID	Response Text
5	Objective comparison or assessment of instructional technique at the different levels.
10	Clearly defined standards for each graded item such that the numbers mean the same thing (roughly) from instructor to instructor.
30	In a fleet squadron, EPs are only practiced on Natops checks and in the EP sim (2801). It does not need to be on the ATF. There should be more items on each ATF for those specific flights/training requirements. The ATFs are comprised of a majority of general graded items and only a couple flight specific items. This forces an IP to try and summarize the stage specific issues in a small area. This should be reversed with a majority of graded items specific to the stage and a few general items.
41	-ATF does not highlight trends well if a PUI is showing improvement or consistent weakness in a particular area. For a particular PUI a graded item might meet a standard of "2" for various flights within the stage but improvment or increasing weakness in the area may only be noted in the remarks or additional comments if the instructor has flown multiple flights with the PUI. -Even with ATF writing training, consistency in ATF writing is not standard between different instructors -ATF does not provide a consistent quantitative assessment of the PUI's performance on a particular event.
50	The current ATF did nothing more than include automatic averaging of [numeric] values, on an equal basis. For one thing, I would argue that the different items be weighted to reflect the actual importance of individual items. Additionally, the IP should be able to more easily indicate, without modifying numerical entries, that a PUI needs additional training. There are times when a PUI can satisfactorily perform all of the existing checklist items on a ATF, but be in need of additional training. The ATF needs to be completely redone. Throw out all of the items and notions that have carried the same ATF for years, and completely redesign it, please. The first, and most important item should be whether or not a PUI needs additional training; it shouldn't be at the bottom of the ATF. Identify it up front, then allow the rest of the ATF to tell why. Additionally, I never cared about 'Use of Checklists' being on an ATF...by the time a Lt gets to the fleet he'she [expletive] better know how to read and execute checklist items, or they shouldn't have made it out of the FRS. I could rant, but I won't. Suffice it to say that our entire concept of the ATF needs to be redesigned.

Table 12. Participant responses on what critical items are currently missing on ATFs

These responses provide some information regarding potential improvements to the ATF. Response number 10 also indicated that the T&R does not clearly describe performance standards, which seems to match the text response provided. This highlights potential conflict between the ATF and what is enumerated in the T&R manual. Participant 30 suggests that more items for each specific flight be graded instead of the

general items that are currently found on the ATF. This may also be a call for merely more flight-specific items while maintaining the current list of standard items on the ATF. Participant 41 believes there must be better trend indication on the ATF, and asserts that ATFs do not provide “consistent quantitative assessment”. The inconsistency is likely due to the large amount of subjectivity and variation in each event. This response also points out that the subjectivity from instructor to instructor is inherent, and causes difficulty in discerning when trends exist. The text associated with response 50 points out that the overall numerical score, in that participant’s opinion, is a poor metric by which to measure performance. Participant 50 also ranked the overall ATF score as a three on the one through seven scale, with one being not important at all, four being neutral, and seven being extremely important in question 16 of the survey. The participant recommends the creation of a weighting scheme to reflect an overall score that is indicative of the critical items.

The final question of the survey asked participants to describe a tool they would like to have at their disposal to aid them in the assessment of PUI performance. The full set of responses can be found in Appendix D. The responses to this question were transferred to a ‘.txt’ file. A simple Java program designed to conduct a word count on a text file written by Dr. Arnold Buss of the Modeling, Virtual Environments, and Simulation Department at NPS for the CS2173 Java as a Second Language course was used to aid in the analysis of the responses (see Appendix E). The program conducts a count of unique words found in the text file searched; however it does not discriminate between derivations of the same word. For example “word” and “words” are considered distinct and counted individually. Despite the very coarse word count provided by the program it was useful in aiding in the identification of themes found among the responses. Two such themes were the mention of subjectivity in evaluation and the desire to see some comparative ability. The word “compare” or some derivative was used 13times and “subjectivity” or a derivative was used nine times. While these counts don’t mean anything when not put in context, they are an indication that these are common issues and interests across the survey participants. These terms were used across five and six responses respectively. The responses that mention subjectivity and comparisons call

for a tool that can show comparison of PUIs across the whole population of PUIs both within a single squadron and across T/M/S in an objective manner. The responses also state that the subjectivity cannot ever fully be removed from the process, which has also been asserted in this thesis.

Another theme expressed in the responses is the idea of inter-rater reliability. Response 11 states the following: “Some type of system similar to FITREP grading average based on the Instructor's average. The problem with the current system is that there is an assumption that all IPs grade the same.” This type of system would provide instructors to view PUIs’ performance through the lens of the IP assigning the grades. It allows the person assessing a PUI via the ATF to better understand the grading profile of the instructor who wrote the report. By comparing the PUI’s grade on a particular item to the average achieved by all PUIs that flew that event with that instructor, those making decisions would be able to judge the PUI by the quality of their score received by a specific instructor. This method does have some short comings, including the fact that this still only allows subjective comparisons across different instructors. This could be overcome by using the magnitude of difference between the PUI’s score and the standard deviation across all PUIs on that event for that instructor. Budrejko (2009) offered several recommendations to standardize the instructor cadre, including inter-rater-reliability measures to provide quantitative measure of success. This method could also be instructive to decision makers if events were weighted in some way so that the overall score could have some overall performance meaning.

Finally, response 32 indicated that methods for making it more clear to decision makers reading ATFs have been attempted within certain units. This response cites the creation of expected and threshold values for each event. This system provides a minimum threshold that a PUI should meet and an expected or typical score for that event. If the threshold is not met a discussion is held among instructors and other leadership as to why the PUI did not meet what they determined as a minimum acceptable score. The survey response did not indicate how the unit arrived at the threshold and expected values for each event. One possible method was to discuss the criteria established on the ATF among instructors and leadership, relate that criteria to the

training model outlined in the T&R manual then come to a consensus on what the group considered a reasonable option. While this may be useful in highlighting when PUIs have difficulty meeting the expectation, it does not identify why they didn't meet it, nor has it been developed with the full instructional system in mind.

3. Comparing Analysis of ATF Data and Survey Results

The data collected and analyzed from both ATFs and the survey must be looked at collectively to synthesize a model for a decision support system. The first items of analysis that can be compared to each other are the ordered plot individual graded ATF items and the order of importance based on the survey results (see Table 13).

RANK OF ATF ITEM AVERAGES	RANK OF ITEM IMPORTANCE
1. Discuss Items	1. Situational Awareness
2. Checklist	2. Headwork
3. Airwork	3. Remarks
4. Situational Awareness	4. CRM
5. Headwork	5. Training Mission Specific Items
6. Communication	6. Discuss Items
7. Mission Planning	7. Communication
8. Emergency Procedures	8. Additional Comments
9. CRM	9. Mission Planning
10. Brief/Debrief	10. Airwork
	11. Brief/Debrief
	12. Emergency Procedures
	13. Checklists
	14. Overall Grade

Table 13. Side by side comparison of ATF item grade average and rank of item importance from survey results

The rank of ATF item importance contains additional metrics that do not include graded metrics. Although training mission specific items receive numerical scores, they were not analyzed in this research because of their specific nature to the individual training event being conducted and are not common across all events. To simplify the comparison for the purposes of illustration the dissimilar items ranked in the survey are removed in Table 14.

RANK OF ATF ITEM AVERAGES	RANK OF ITEM IMPORTANCE
1. Discuss Items	1. Situational Awareness
2. Checklist	2. Headwork
3. Airwork	3. CRM
4. Situational Awareness	4. Discuss Items
5. Headwork	5. Communication
6. Communication	6. Mission Planning
7. Mission Planning	7. Airwork
8. Emergency Procedures	8. Brief/Debrief
9. CRM	9. Emergency Procedures
10. Brief/Debrief	10. Checklists

Table 14. Side by side comparison of ATF item grade average and rank of item importance from survey results with non-standard graded items and non-numerical standard items removed from rank of item performance column

First, we notice that situational awareness and headwork are grouped in both lists as pairs as are communication and mission planning. This may be due to instructors grading these items in a similar fashion when determining how PUIs perform. An interesting point here is that situational awareness, headwork, and communication are related to performance in the aircraft, while mission planning is generally a pre-flight consideration. We also notice that checklists rank second highest by average but are considered the least important by IPs who participated in the survey. This matches the comment found in survey response 50 that PUIs should be familiar with and capable of executing checklists by the time they reach the operational squadron. The same might be said for emergency procedures, with regards to IP expectation of PUI performance. Very rarely in the operational squadron are emergency procedures drilled during the syllabi, except when done in conjunction with required recurring events. Further complicating the analysis is that there could be interactions between graded items. For example, if a PUI does poor mission planning, and as a result receives a poor mark on that item on the ATF, he or she might also have poor situational awareness or headwork during the execution of the same training event. The next observation is that some items expected to be mastered by a PUI by the time they begin training beyond the Core Skill Introductory Phase, namely checklists and airwork, receive high marks, but are of low importance. As a result

these items inflate overall averages and may result in misinformed readers of an ATF when the total average score is utilized as a measure of performance. This may be another reason as to why overall scores are considered least important among those surveyed.

C. TOOL DESIGN AND MODELING

The intent of this tool is to improve the design of the instructional framework for Marine aviation. Current efforts aimed at improving the tools in use to evaluate readiness address issues that include providing electronic data warehouses and an electronic means to complete ATFs. These efforts are outlined in a contract solicitation that includes a performance work statement that outlines the expansion of the MSHARP system (Commanding General Regional Contracting Office National Capital Region, 2014). This solicitation does not include any capability for built in analysis for ATF data. The model described in this section could be fully developed and integrated into the MSHARP interface to provide the improved resolution for training evaluation.

The previous section detailed the analysis conducted on information collected from ATFs and the survey conducted to solicit instructor pilots' opinions on the current ATF and identify those aspects they consider important. This information was then used to inform the design of an output prototype that provides a comparative and quantitative assessment tool.

1. Design of an Item Weighting Scheme

Based on the analysis we will assume for this development that the overall average of pilot performance is non-instructive for decision makers. It can be manipulated to artificially inflate or deflate grades to achieve a particular overall score by the instructor. This requires a new method for calculating an overall score, which can be done by creating a new model for scaling individual item scores to provide an overall score that is more instructive. The new model also must differ from the calculation of the NSS because the NSS is utilized under the assumption that an SNA has completed the full course of training, accounting for all phases of training prior to designation as a naval aviator (Naval Air Training Command, 2007, Appendix E). The survey results provide a

ranking of importance of each item graded on the ATF. This importance was translated into a weighting scheme that reflects the importance level judged by instructors on each scored metric (see Table 15).

Proposed Weighting Scheme for Overall Score Calculation (excludes mission-specific items)		
Tier 1	**Situational Awareness**	17.5%
	Headwork/Decision Making	17.5%
Tier 2	**Crew Resource Management**	10%
	Discuss Items	10%
	Communication	10%
	Mission Planning	10%
	Airwork	10%
Tier 3	**Brief/Debrief**	5%
	Emergency Procedures	5%
	Checklist	5%

Table 15. Proposed weighting scheme for ATF graded items excluding mission-specific items

The proposed weighting does not account for flight-specific items because these items were not recorded in the data-set used in the analysis. These items could easily be incorporated with minor adjustments. Table 16 offers a weighting scheme that includes mission specific items.

Proposed Weighting Scheme for Overall Score Calculation (includes mission-specific items)		
Tier 1	**Situational Awareness**	17.5%
	Headwork/Decision Making	17.5%
Tier 2	**Crew Resource Management**	8%
	Training Mission-Specifics	8%
	Discuss Items	8%
	Communication	8%
	Mission Planning	8%
	Airwork	8%
Tier 3	**Brief/Debrief**	6%
Tier 4	**Emergency Procedures**	5.5%
	Checklist	5.5%

Table 16. Proposed weighting scheme for ATF graded items including mission-specific items

To demonstrate the differences between the weighted and non-weighted averages the event averages were calculated, and can be seen in Table 17.

Event	AVG	Std Dev	Weighted AVG	Std Dev
FAM1111	2.89	0.324	2.89	0.379
FAM1117	2.95	0.414	2.98	0.484
FORM1303	3.08	0.278	3.08	0.369
SWD1602	3.03	0.288	3.06	0.401
SWD1605	3.12	0.206	3.10	0.284
TERF2100	2.22	0.349	2.21	0.365
REC2300	2.49	0.408	2.48	0.413
SWD2602	2.18	0.273	2.15	0.279
SWD2604	2.28	0.334	2.27	0.379
SWD2607	2.42	0.386	2.43	0.397
ANSQ2705	2.38	0.325	2.37	0.345
ESC3103	2.59	0.273	2.56	0.304
CAS3303	2.49	0.367	2.51	0.380
AI3306	2.41	0.364	2.43	0.343
AHC6398	2.93	0.199	2.89	0.186

Table 17. Comparison of non-weighted and weighted averages and standard deviations

The weighted averages have an increased standard deviation, which gives the decision maker a greater resolution on stratification of PUI performance. There are only two events that when scored with the weighting scheme applied had a smaller standard deviation. We suspect that with a larger sample, this would likely not be the case. In addition, the weighted averages values that are at the extremes inform the observer that a PUI has done poorly or well on the items that are considered most important. The weighted averages are informative, but they still fall short of providing a full picture of PUI performance. This value does not provide information on how trainees are performing relative to their instructors' grading tendencies.

2. Design of Graphical Component Prototype

In order to provide decision makers with comparative information a graphical representation of PUI performance based on scores was constructed. Based on analysis of

performance data provided from ATFs and the results of our survey, the items that will be useful in assessment are PUI overall averages compared to the population of PUIs. This includes overall averages, averages grouped by event, and averages grouped by item. Instructors also expressed interest in having the capability to understand how PUIs perform with specific instructors on an event compared to other trainees who have flown with that instructor. To gain the full understanding of a trainee's performance a decision maker must observe these metrics simultaneously. A basic depiction of this layout is seen in Figure 26.

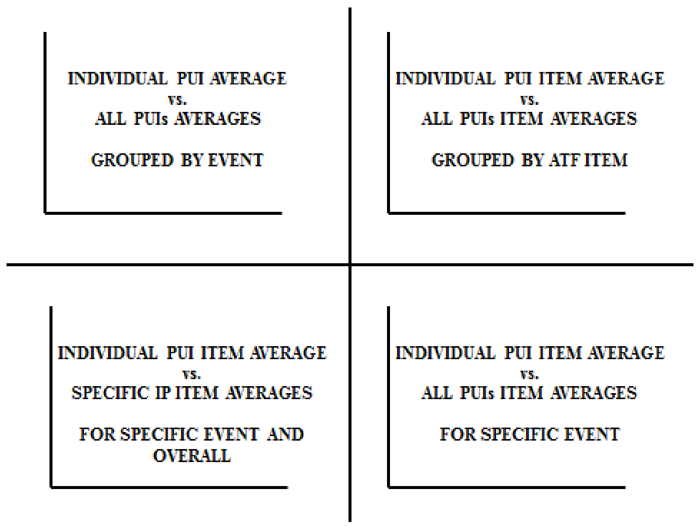


Figure 26. Simultaneous viewing of ATF scores for a specific trainee

This layout provides a snapshot a PUI's performance over the course of the syllabus, as well as comparative charts that give an indication of what the instructor's grading profile for that specific graded event looks like. The image seen in Figure 27 utilizes AH pilot sample data to provide performance information visually to the ATF reviewer or decision maker.

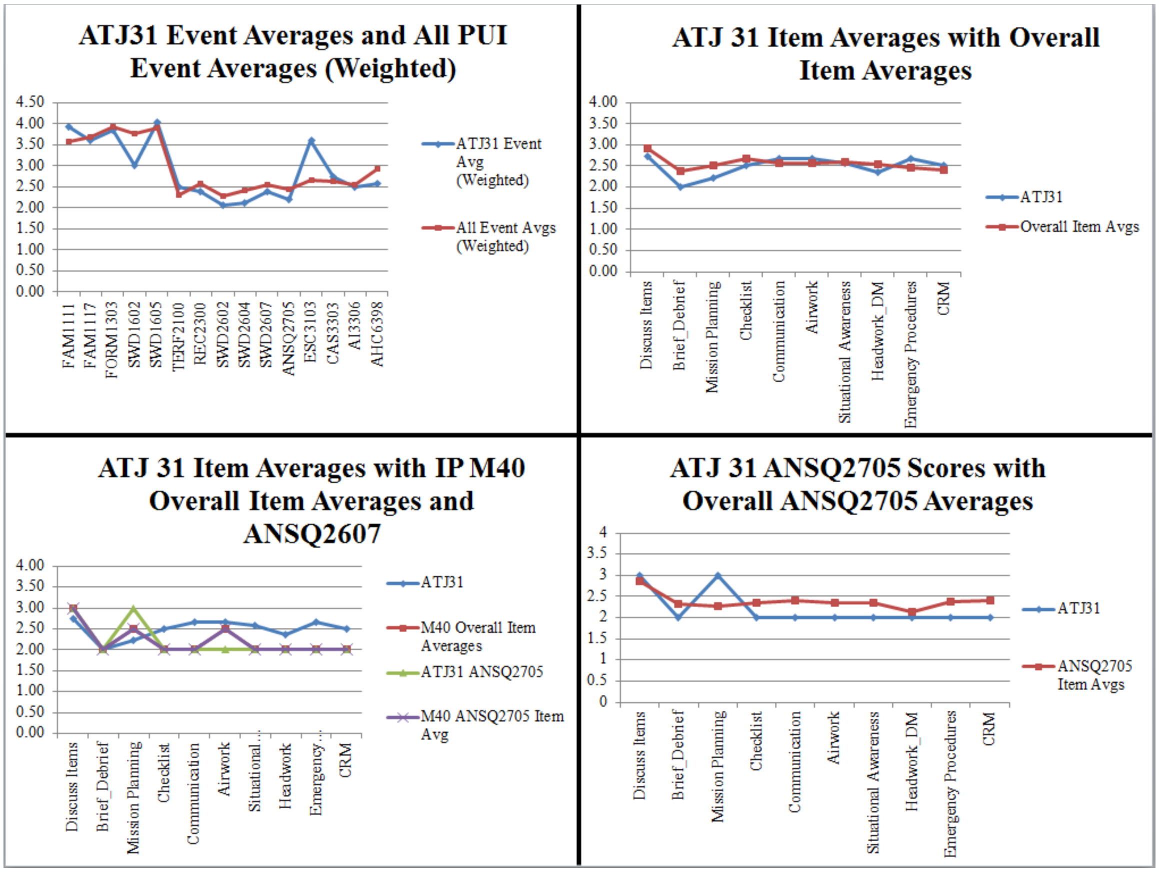


Figure 27. Comparative performance output for an individual and specific event

This display is rudimentary and for illustrative purposes. To improve the meaning of the display for the decision maker, each chart would include some interactive capabilities. These features would include mouse-over capability to display values for individual points, the ability to adjust the size of the image, and the option to display standard deviation for each measure if so desired. Were performance thresholds to be developed, they also could be plotted and displayed. Should a PUI fall below those thresholds or exceed them, simply shading the quadrant containing specific event comparisons red or green would provide immediate indication that PUIs are failing or completing events. A chart of this type has been tested for use at the FRS for H-1 aircraft (see Figure 28).

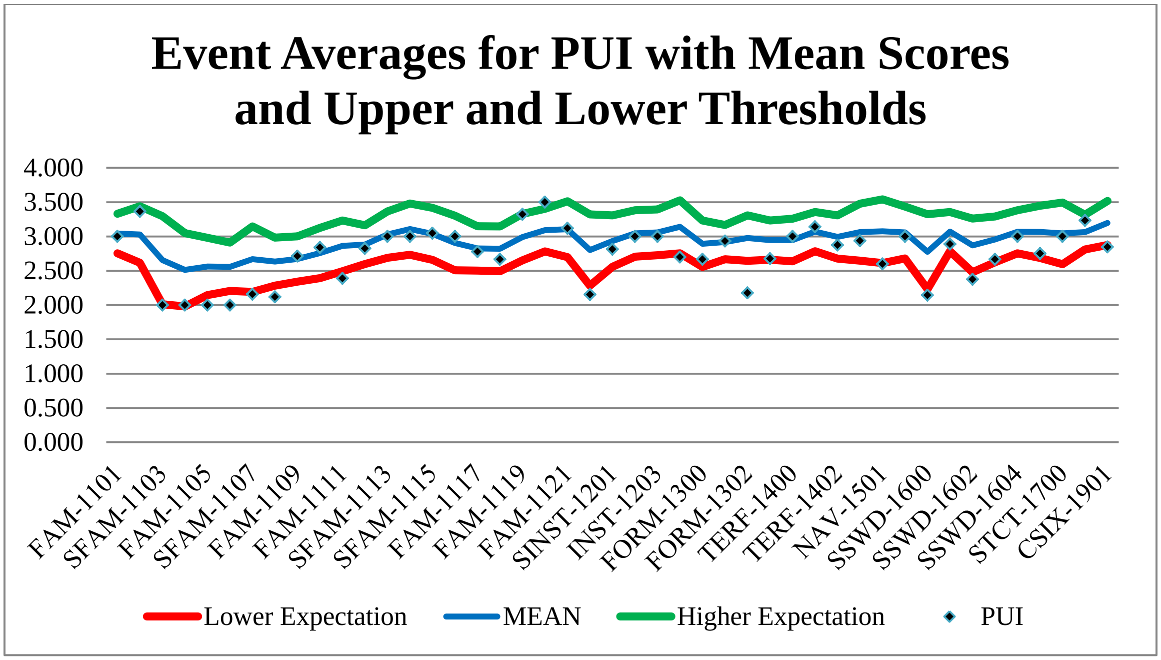


Figure 28. Plot of PUI event scores at FRS with all PUI mean scores and upper and lower expectations (after Marine Light Attack Training Squadron 303 Operations Department, 2013)

The information displayed in Figure 27 could be customized further by the individual attempting to assess trainees by use of information provided on ATFs by allowing the user to choose what exactly the wish to compare. For example, if the evaluator chose to look at a PUI's last flight they could select the PUI, and the event from a menu, and they would be presented with the above graphic and access to the full ATF for the specific event undergoing review. An example of such a selection interface is shown in Figure 29.

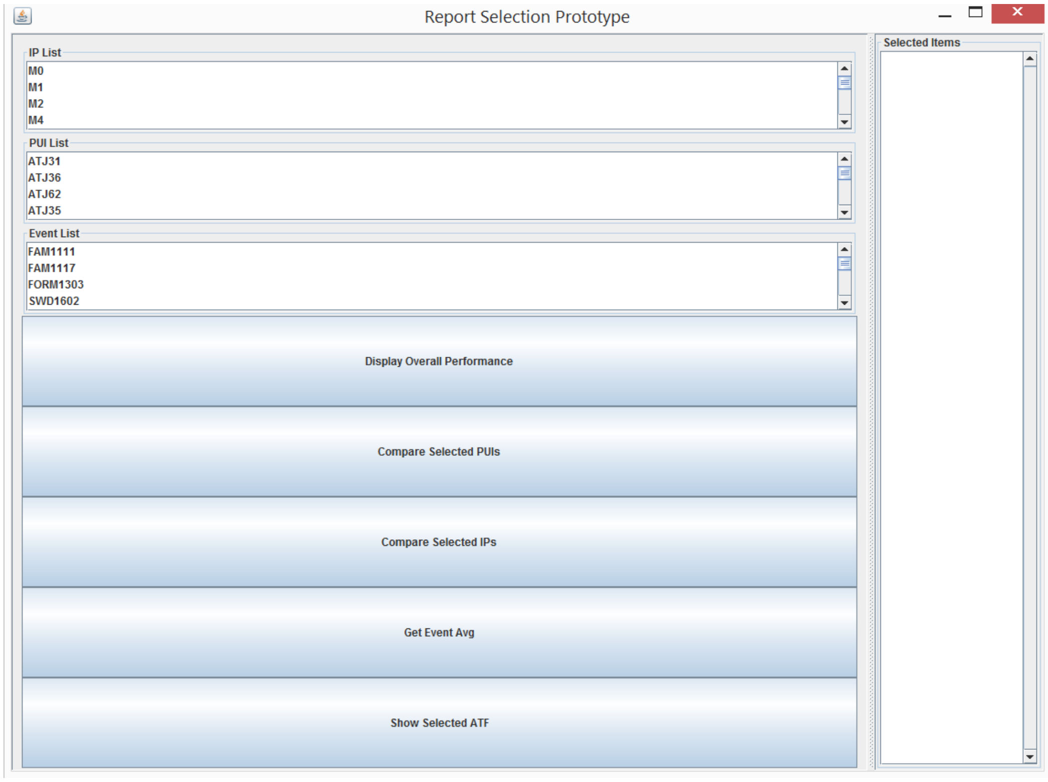


Figure 29. Example Report Selection Interface

Interaction with such an interface would allow decision makers to begin by selecting an IP, a PUI or an event. Once one of these items is selected, the remaining lists would automatically update to reflect possible relevant selections. In this manner invalid combinations would not be an option for displaying reports. To further expand the capability of the interface, one could integrate the ability of senior squadron leaders who are required to review and sign all ATFs to electronically sign the document following receipt of the comparative report. The new capability that is gained in the above model is the availability to review and compare IP grading tendencies. This capability enhances the meaning of scores received by trainees. By visualizing the performance of PUIs, decision makers are presented with information that indicates the potential need for training intervention in a specific area. These interventions could make use of existing simulation technology, to enhance a trainee’s performance of a particular skill-set.

3. Integration of the Proposed Tool

MSHARP currently provides users access through a web interface with selectable readiness reports and ability to export raw data in spreadsheet format. As discussed earlier, the existing interface does not possess this ability for ATF score data. Nor is ATF completion data linked to event completion for readiness reporting. By integrating the proposed capabilities into the MSHARP interface, unit-level evaluation processes would be made more efficient. The most recent solicitation for enhancements to MSHARP provide for data warehousing, options for making the web interface compatible with mobile devices, and electronic record keeping of ATFs (Commanding General Regional Contracting Office National Capital Region, 2014). If the analysis tools proposed by this research are made available through the MSHARP interface, feedback to PUIs will become more concrete. The tools here could become available based on a hierarchy of permissions through which all participants in the instructional system may be involved more heavily in the ATS. For example, senior decision makers have the ability to see all individual aviators within the unit, while individuals may view only their own performance against the averages of other individuals. The current unit work flow is displayed in Figure 30 at the macro level. The diagram does not depict the requirement of the IP to fill out the ATF independently of MSHARP. The reader should also notice the lack of formalized comparison of the PUI's performance to other individuals. This means that IPs must rely on their previous experience of assigning scores or having been assigned scores by others when accessing through their own syllabi.

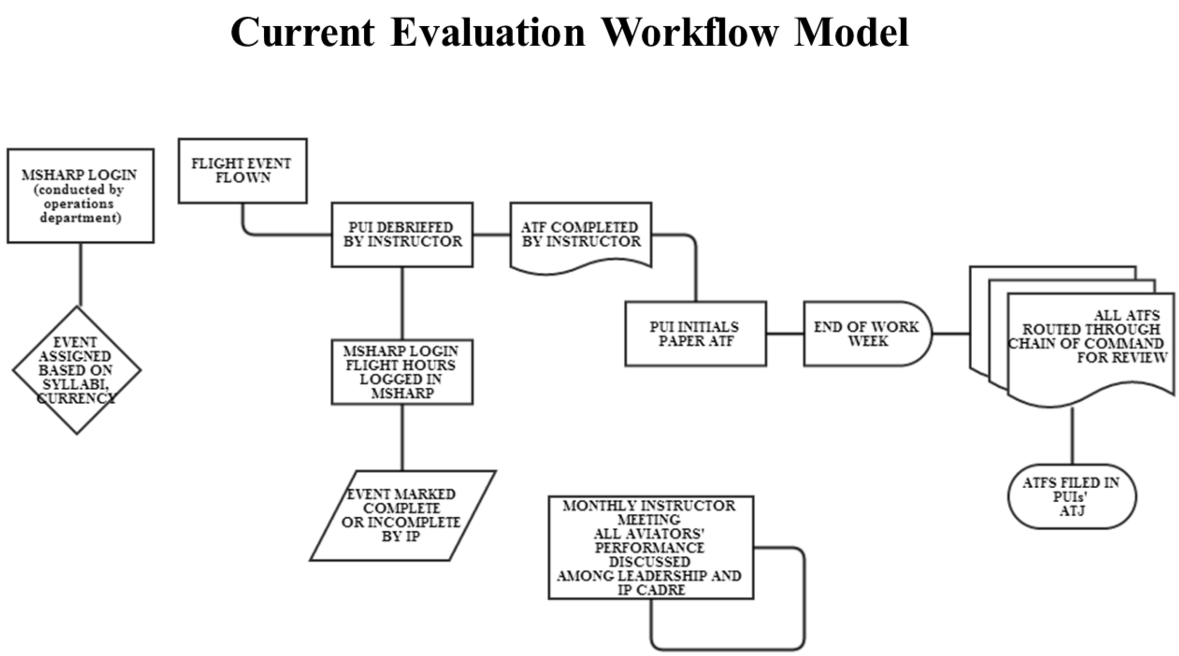


Figure 30. Current unit-level work flow for PUI assessment

This process could be streamlined by integrating the proposed tool into the evaluation process. It would also provide for relevant comparisons that would improve decision makers' ability to initiate training interventions when they deemed necessary, because they are made aware of deficiencies in a more meaningful way. A model of the process is shown in Figure 31.

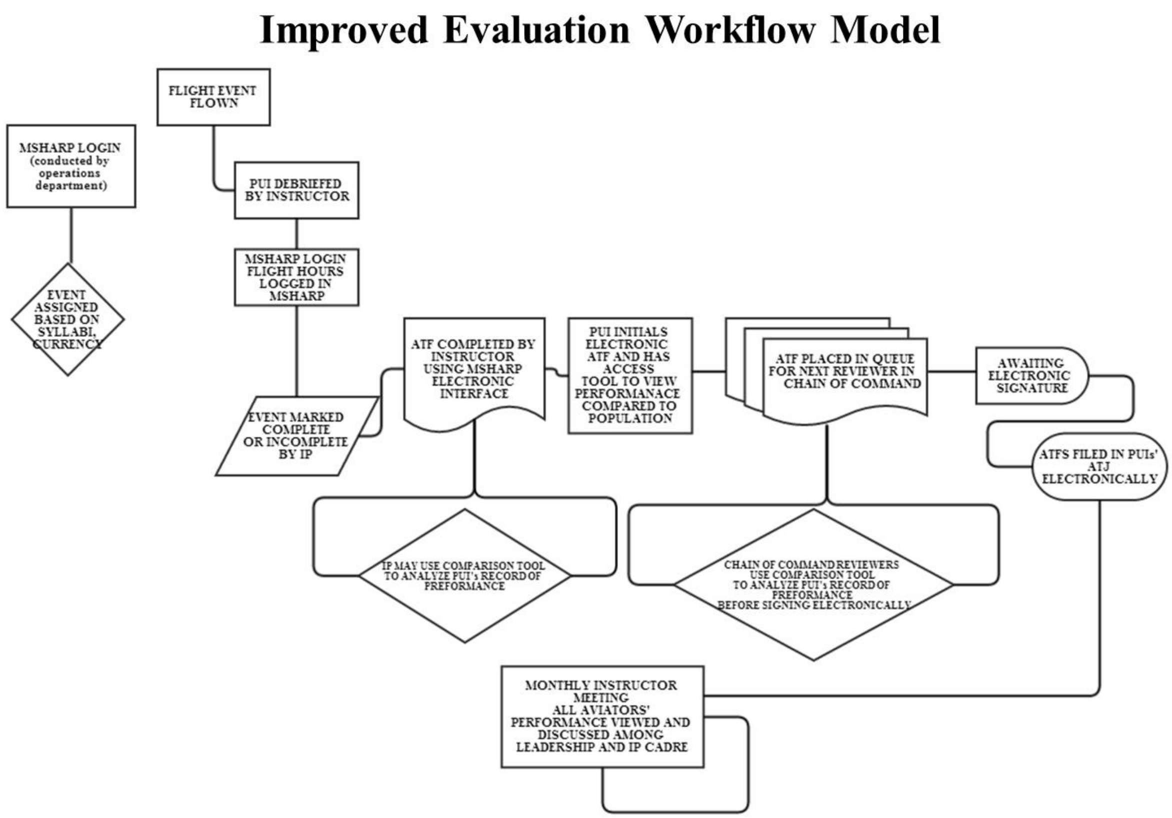


Figure 31. Improved unit-level work flow for PUI assessment

In this model feedback is provided to PUIs through electronic access to their records and comparison tools that show their performance compared to the population of aviators who have also completed the same events. The reader will also notice that reviewers immediately have access to the completed training documentation once it is reviewed by the PUI. In addition leadership will use the comparison tools described earlier to better understand the PUI performance. This allows the decision maker to attend instructor meetings armed with new information on the performance within the unit. During instructor meetings the proposed tool would also be available to the group to focus discussion of performance on trends and methods to remedy deficiencies as well as recognize exceptional performance. Through this integration the instructional system is improved, and is closer to reflecting a complete instructional model.

D. CHAPTER IV SUMMARY

The tools described in this chapter enhance the instructional system by providing information that was previously extremely impractical to derive from the ATF system as it is currently implemented. As evidenced in the survey responses detailing how specific units have established thresholds of success and failure, as well as the disparity of items considered important found on evaluation forms, the ATS is failing to standardize the evaluation procedures across Marine aviation. Reliance on paper documents to quantify performance fails to maximize the usage of information available. Previously, it may have required several poor flights by a trainee for leadership to recognize a trend. With the usage of the data that is created by the instructional system and presented via the proposed tools, earlier recognition of marginal performance as well as exceptional performance is feasible.

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The portion of the current instructional system used to assess the performance of Marine aviators is incomplete and does not provide an efficient and effective means to assess and compare the performance of individual aviators within a unit or across multiple units. The amount of data created by ATFs within a squadron does provide a baseline with which to assess aviator performance; however, the data is not formatted in a manner that provides decision makers with a snapshot of pilot overall performance, nor does it provide a means to visualize trend information on trainees. After analyzing a sample of operational training performance, we were able to create an initial prototype of a system capable of providing a visual display that conveys trainees' comparative performance within the instructional system. The prototype is also capable of providing information regarding instructor performance and trends. The inclusion of tools as described in this research in the instructional system would provide a feasible method to evaluate the instructional system that currently does not exist.

The current formulation of criterion-based scoring and of the overall grade achieved by trainees does not provide sufficient information to derive predictive measures for future performance. Although one might expect grading scores to begin at a baseline level and progressively improve through a specific phase of training, this pattern is not exhibited in the data analyzed. Due to large numbers of missing values found in the ATF data-set, correlations between success and failures were not found. This research did identify that there is a significant difference between scoring at the FRS and within operational squadrons, which is likely due to slightly different grading criteria. We also found that there are significant differences between scores of pilots of different types of aircraft. Based on the singular unsatisfactory flight found in the data analyzed, further research must be conducted to understand identifying score- and comment-factors that are indicative of future performance.

B. RECOMMENDATIONS

The analysis of training data and the survey conducted provided a basis for which to provide a series of recommendations that can guide the future development of a training diagnostic tool for Marine aviation and adjustments to the instructional system used by the ATS.

The first recommendation is to realize electronic collection and storage of performance data contained on ATFs, and to have that data accessible in a fashion which supports the ability to conduct analysis. Efforts of this nature are currently underway but do not include automated analysis capability. This means that units will require manual manipulation of the data to glean meaningful performance information. By storing this data electronically, a host of new capabilities become available to decision makers. Despite being electronically stored, this data may not see immediate use without automated analysis capability. This is underscored by the fact that units have attempted to develop and implement thresholds and manually enter and record data into excel spreadsheets from paper reports that have been generated.

The second recommendation is that further development of the criterion-based scoring system be conducted. The current implementation requires that instructors apply a rubric to items that do not match the skills and activities to be performed. For example, it is very difficult to standardize what correct, efficient, skillful and without hesitation is in regard to situational awareness. When a PUI receives a score of four, the criterion reference is that they require no further instruction. This may be true for a very specific procedure-based task; however it is unlikely that any trainee ever demonstrates perfect crew resource management or headwork and cannot improve. While the application of the current rubric may be acceptable to these items, better descriptions of what these scores mean in context to these items may improve the information contained in these records. Coinciding with reformation of criterion descriptors is the implementation of a weighting scheme placed on scored items. This weighting scheme, if known and understood by instructors and decision makers, can give meaning to an overall score. The current formulation of the score provides only an indication of a macro view of

performance, where a weighted score can provide information on whether a PUI is performing adequately on items deemed most important by instructors and leadership.

A third recommendation is to refactor the ATF to include only items that instructors deem necessary to provide a score for and to remove from the list items that are expected to be completed proficiently, such as checklists. Emergency procedures are rarely assigned a score by instructors. This is because the emphasis of the majority of training flights in an operational squadron is on tactical mission skills. This can be handled by incorporating emergency procedure review into each event by focusing on one or two during a training event and providing a pass or fail block on the ATF. Emergency procedures are expected to be known at all times by designated aviators and this would build an additional ability to build proficiency through knowledge of procedures and situations in which to apply them. Should a trainee fail to have sufficient knowledge it would be marked as such and the instructor would have the ability to comment the deficiency as necessary.

The final recommendation is to integrate the proposed tool into the currently existing MSHARP interface and into the electronic ATF generating component within MSHARP once it is established. By making the analysis tool an integral part of the ATF writing and reviewing process, it is integrate into the instructional system and ensures that it is utilized by those that write ATFs and review performance of trainees. If it were to be a stand-alone capability, it may not be utilized to the greatest extent possible.

1. Future Research Efforts

Two main areas of research can be pursued from the work completed in this thesis. The first is the matter of gaining further understanding of how and why instructors assign particular scores to trainees, what those scores reveal about the individual, and what they reveal about groups of trainees when analyzed. The second is the continued design, development, and integration of an automated analysis tool that provides leadership with training intervention decision support. Both of these issues for research will benefit greatly from the planned digitization of records that will provide access to the data to support these goals.

Once records are fully available in electronic format and a sufficient database of ATF records is created analysis should be conducted on those records in a similar fashion to what was done in this research. It is recommended that future research utilize training data from additional units and across dissimilar types of squadrons. The transition to electronic records will also provide the ability to access written comments and discussion on pilot performance. It is also recommended that semantic analysis be conducted to better understand how comments recorded on ATFs provide indicators of future performance. These comments may inform the understanding of what factors influence the scores being assigned.

Future efforts in development of a tool that can support decisions regarding training interventions should focus on user studies that evaluate how the tool can influence leadership to provide training interventions. These research efforts will require working prototypes that can be inserted into the instructional system to conduct user studies on ease of use and training outcomes for trainees and instructors who had regular access and use to such a tool. The development of this tool must also address the evaluation of instructor cadre, which is absent from the current instructional system. These tools should be made accessible to units when conducting review boards and instructor meetings meant to assess progression of PUIs through their respective syllabi. In this way comparisons can be made between units training aviators with and without the system in question.

In the end, we hope that this research and these recommendations result in a fully developed instructional system, and provide a model for a framework that can be utilized across all training domains in the development of instructors and trainees in their respective warfighting domains.

APPENDIX A. SAMPLE AVIATION TRAINING FORM

WTO-5203

WTO-5203	HRS: 3.9	2 AH-1W A D	CREW: WTO/IUT
IUT: Capt NAME OF PUI	IP: Capt NAME OF INSTR		DATE: 16 FEB 2012

Goal: RS - Repeat SWTO-5201 in the aircraft with emphasis on instructional techniques and tactics standardization.

Requirement: See AH-1W T&R

Performance Standards: See AH-1W T&R

Prerequisite: WTO 5200-5202

Ordnance: (7) 2.75" rockets, (300) 20mm, (1) captive PGM, (30) flares

Mission Profile / WX: NFG-2507N-NFG / 05012KTS FEW055 10SM

1. GENERAL	DND	UN	1	2	3	4	REMARKS
Discussion Items					X		
Brief / Debrief					X		Well done. Brief tailored specifically to audience for CAS/BCWD. Included all required information to execute flight and amplifying instruction for PUI
Mission Planning					X		
Checklist					X		
Communication					X		
Airwork					X		
Situational Awareness					X		High throughout TOS and flight overall, with mutiple final controllers (Ground and FAC(A)), and a high tempo in the objective area.
Headwork					X		
Emergency Procedures	X						
CRM					X		Managed appropriately. Remember with a new pilot in the rear seat the requirement to ensure they can put the A/C where it needs to be at the appropriate time goes up exponentially. With a PUI expect to have to provide more guidance than you would flying with a peer or more senior pilot - and most importantly, no matter who it is know when it is time to speak up - which you did today.

2. PERFORMANCE	DND	UN	1	2	3	4	REMARKS
Rocket Delivery					X		Good instruction technique and error correction for rear seat pilot
20mm Delivery	X						Appropriate CRM and cockpit management during attacks despite gun not firing. Good troubleshooting steps while still managing external comms and keeping tempo up in
Error Recognition & Correction					X		
Instructional Technique					X		

Grade: 3.00

DND – Not applicable or not observed.
UNSAT* - Unsafe or complete lack of ability or knowledge. Requires substantial input from IP for safe execution and/or mission accomplishment.
1 – Safe but limited proficiency. Requires frequent input from the IP.
2 – Correct. Recognizes and corrects errors. Requires occasional input from the IP.
3 – Correct, efficient, skillful, and without hesitation. Requires minimal input from the IP.
4 – Unusual high degree of ability. No further instruction required.* Mandatory comments for items scored at this level.

3. FLIGHT SUMMARY:	REMARKS
Strong Points:	Solid Brief and High SA throughout
Weak Points:	None noted

PUI: _____ OPSO:_____ DOSS:_____ XO:_____ CO:_____

WTO-5203 (V1)

WTO-5203

Overall Flight:	X	**COMPLETE**		**INCOMPLETE**
		R.A.T.		**UNSATISFACTORY**
Recommendations:	IUT has not completed 5202 yet and has not executed a flight specificall focused on BCWD. Recommend that he be scheduled for 5202 on a dedicated BCWD event in order to be exposed to leading a flight with line numbers and allow IUT to refine error correction, instructor technique, and and instructor demonstration abilities.			

COMPLETE: Declares the pilot under training has demonstrated sufficient grasp of the concepts and skills to proceed to the next training evolution or be designated appropriately.

INCOMPLETE: Describes a training event that is not declared 'Complete' due to circumstances beyond the control of the aircrew. Examples may include, but are not limited to: WX, time constraints, aircraft or simulator maintenance, external support inadequate. 'Incomplete' shall not be used to obscure reporting of a substandard performance.

REQUIRES ADDITIONAL TRAINING (R.A.T.): Not derogatory in nature. The pilot under training has not yet demonstrated sufficient grasp of the required skills and concepts to progress. Instructor remediation recommendations should specifically identify the deficient area(s) for addressing shortcomings in terms of reading assignments, courseware, additional flight, simulator, or other appropriate training. The Instructor assigning a R.A.T. synopsis is responsible for ensuring the recommendation has been endorsed by Squadron leadership and adhered to by the student unless a higher authority intervenes with additional guidance.

UNSATISFACTORY: Identifies a condition where the pilot under training has proven unable to meet performance standards due to a lack of preparation, lack of effort, consistent inability to demonstrate improvement or resistance to instruction. Significant safety of flight incidents that are of a direct result of the pilot under training actions should be considered unsatisfactory. The instructor assigning this event synopsis is responsible for ensuring recommendations for remediation, if applicable, are proposed through the DOSS & Operations Department.

4. ADDITIONAL COMMENTS:

Flight briefed as pure AH section as flight ISO Scorpion fire in 2507N and flown from the F/S. IUT conducted a thorough section brief to execute the time on station safely and effectively. IUT tailored the brief appropriately to the audience to include -2 who was receiving a BCWD X. Appropriate level a detail required was provided for all phases of flight and instruction was offered continuously.

During execution IUT maintained a high level of SA in the objective area working the section for several terminal controllers and managing the responsibilities of being LD and an instructor within the crew. Remember that as an instructor you usually aren't just instructing within your cockpit, but you are also usually leading and managing the section and ensuring the safe operation of the flight, adhering to range regs, and providing appropriate and timely service to the customer if supporting some other element. Today IUT managed all of these items in a efficient manner. Flight executed several attacks to include multiple Type IIIs. During each attack IUT provided appropriate instruction and guidance within cockpit to ensure instructor acting as PUI to successfully deliver ordinance against targets.

Overall an excellent flight which provided the IUT with exposure to managing all of the facets of flight instruction within a busy and challenging objective area. Recommend IUT execute a dedicated BCWD for his 5202 to refine error correction and instructor technique.

Capt NAME OF INSTRUCTOR
CALLSIGN

PUI: _____ OPSO:_____ DOSS:_____ XO:_____ CO:_____

WTO-5203 (V1)

APPENDIX B. INSTRUCTOR PILOT OPINION SURVEY

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% [ ] 100%

Informed Consent

* *IC*

You have been invited to participate in a survey for a research study conducted by the Naval Postgraduate School entitled "A Statistically Based Training Diagnostic Tool for Marine Aviation." The purpose of this study is identify critical assessment metrics and incorporate those items into a training diagnostic tool that can aid instructors and leadership in identifying strengths, weaknesses, and trends among aviators within the squadron. These critical items are being identified through analysis of squadron ATFs and through the responses collected in this survey.

This survey is expected to take 20 minutes to complete. If you choose to participate you will respond providing your opinion on this subject matter. There are no foreseeable risks associated with participating. You will not directly benefit from participating.

Every reasonable effort will be made to ensure that responses remain protected and anonymous. All data obtained through this study will be collected anonymously and stored on a password protected computer at the Naval Postgraduate School and not shared with members outside the research team. Your individual responses may be quoted in the published research.

If you have any questions or comments about the research, or you experience an injury or have questions about any discomforts that you experience while taking part in this study please contact the Principal Investigator, Dr. Sam Buttrey, (831) 656-2595, buttrey@nps.edu. Questions about your rights as a research subject or any other concerns may be addressed to the Naval Postgraduate School IRB Chair, Dr. Larry Shattuck, lgshattu@nps.edu

Do you consent to participate?

Please select at most one answer

- ☐ I consent to participate
- ☐ I do not consent to participate

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

Demographics

* *1* **Select all instructor designations you currently hold:**
Check any that apply

- ☐ BIP
- ☐ TERFI
- ☐ WTO
- ☐ TSI
- ☐ NSI
- ☐ DACMI
- ☐ FAC(A)I
- ☐ FLSE
- ☐ FRSI
- ☐ NSFI

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

Demographics

* **2 How many total military flight hours have you logged in your career thus far?**
Please select at most one answer

- ☐ Fewer than 600
- ☐ Between 600 and 1000
- ☐ Between 1000 and 1500
- ☐ Between 1500 and 2000
- ☐ Greater than 2000

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

* *3*

To what extent do you agree or disagree with the following statement:

"The performance standards in the Training and Readiness Manual for my T/M/S are clearly defined."

	Strongly Disagree 1	2	3	Neutral 4	5	6	Strongly Agree 7
Choose One:	○	○	○	○	○	○	○

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

* *4_14*

Rate the following standard items from the ATF based on importance to you when entering numerical scores and when reading ATFs to assess the performance of a PUI.

	Not Important at All 1	2	3	Neutral 4	5	6	Extremely Important 7
Discussion Items	○	○	○	○	○	○	○
Brief / Debrief	○	○	○	○	○	○	○
Mission Planning	○	○	○	○	○	○	○
Checklists	○	○	○	○	○	○	○
Communication	○	○	○	○	○	○	○
Airwork	○	○	○	○	○	○	○
Situational Awareness	○	○	○	○	○	○	○
Headwork	○	○	○	○	○	○	○
Emergency Procedures	○	○	○	○	○	○	○
CRM	○	○	○	○	○	○	○
Training Mission Specific Items	○	○	○	○	○	○	○

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

* *15* On a scale of 1-7 how important do you consider the "Remarks" section for each graded item on the ATF?

	Not Important at All 1	2	3	Neutral 4	5	6	Extremely Important 7
Choose Importance:	○	○	○	○	○	○	○

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

* *16* On a scale of 1-7 how important do you consider the overall numerical grade on the ATF?

	Not Important at All 1	2	3	Neutral 4	5	6	Extremely Important 7
Choose Importance:	○	○	○	○	○	○	○

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

* *17* On a scale of 1-7 how important do you consider the "Additional Comments" section for an ATF?

	Not Important at All 1	2	3	Neutral 4	5	6	Extremely Important 7
Choose Importance:	○	○	○	○	○	○	○

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

* *18*

To what extent do you agree or disagree with the following statement:

"A 'Required Additional Training' grade, in your opinion, is not derogatory in nature."

	Strongly Disagree 1	2	3	Neutral 4	5	6	Strongly Agree 7
Choose One:	○	○	○	○	○	○	○

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

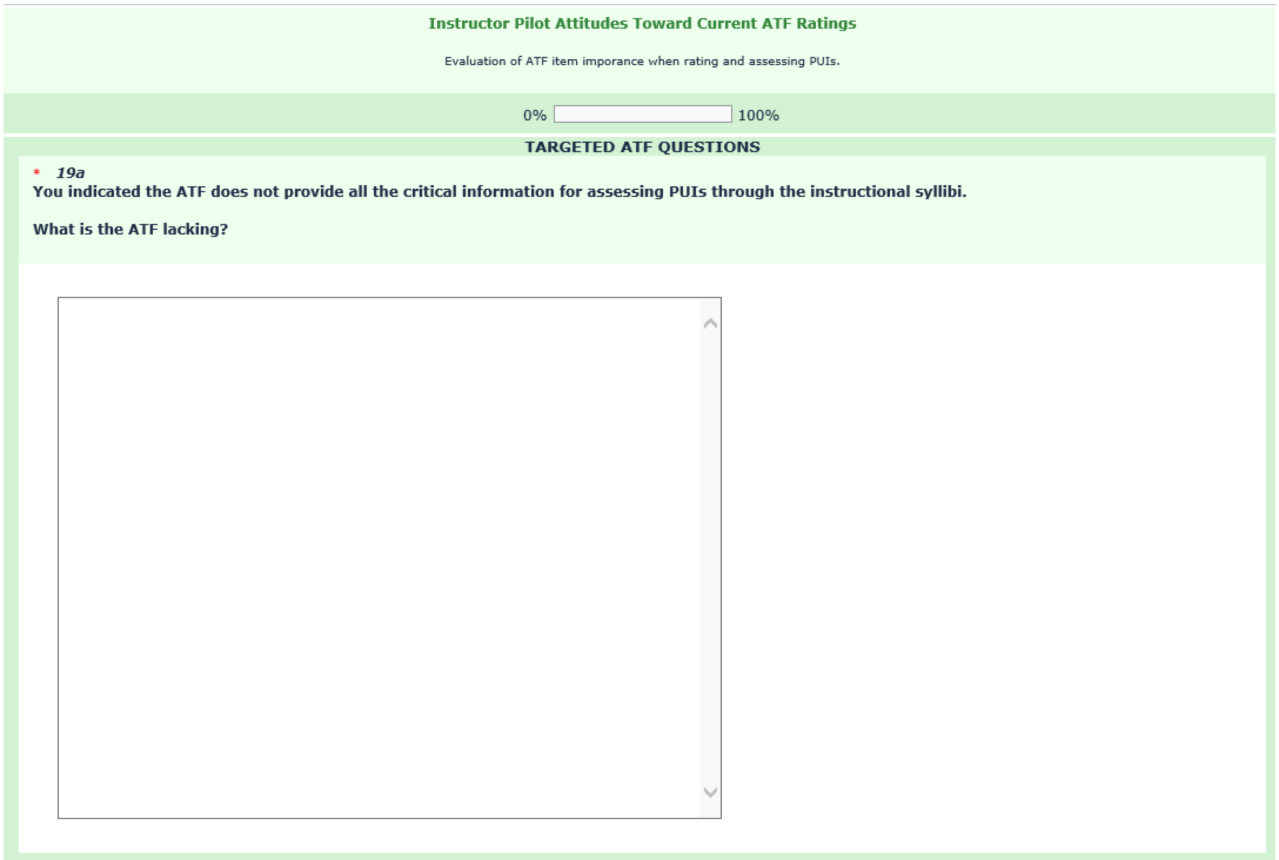
TARGETED ATF QUESTIONS

* *19*

To what extent do you agree or disagree with the following statement:

"Overall, the current layout and form of the ATF provide the necessary critical information for assessing the progression of the PUIs through the instructional syllabi"

	Strongly Disagree 1	2	3	Neutral 4	5	6	Strongly Agree 7
Choose One:	○	○	○	○	○	○	○


Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

* *19a*
You indicated the ATF does not provide all the critical information for assessing PUIs through the instructional syllibi.

What is the ATF lacking?

*19a Asked only if selected one through three on question 19

Instructor Pilot Attitudes Toward Current ATF Ratings

Evaluation of ATF item imporance when rating and assessing PUIs.

0% 100%

TARGETED ATF QUESTIONS

*** *20* If you were to have access to a tool which was meant to aid you as an instructor in assessing the performance of a PUI or group of PUIs, what capabilities would you like it to possess?**

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. INSTRUCTOR PILOT RECRUITMENT EMAIL

SUBJECT: Survey of MAG-39 Instructor Pilots - IF NOT AN INSTRUCTOR DISREGARD

Ladies and Gentlemen,

If you currently hold any instructor designations you have been invited to participate in a survey for a research study entitled "A Statistically Based Training Diagnostic Tool for Marine Aviation." The purpose of this study is identify critical assessment metrics and incorporate those items into a training diagnostic tool that can aid instructors and leadership in identifying strengths, weaknesses, and trends among aviators within the squadron. These critical items are being identified through analysis of squadron ATFs and through the responses collected in this survey.

The survey is only 20 questions long and should take less than 20 minutes of your time to fill out.

Please take a few moments of your time to fill out the survey and potentially help create a useful and meaningful training diagnostic tool for Marine Aviation.

The following link will take you to the survey website: https://survey.nps.edu/546248/lang-en

Participation in this survey is strictly voluntary, and should you have any questions or comments about the research, or you experience an injury or have questions about any discomforts that you experience while taking part in this study please contact the Principal Investigator, Dr. Sam Buttrey, (831) 656-2595, buttrey@nps.edu. Questions about your rights as a research subject or any other concerns may be addressed to the Navy Postgraduate School IRB Chair, Dr. Larry Shattuck, lgshattu@nps.edu

Thank you for your time in advance.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. TABLE OF RESPONSES TO SURVEY QUESTION 20

Response ID	Response Text
2	The most important capability in assessing the performance of PUIs is the ability to input meaningful comments regarding the student's cognitive, affective, and psychomotor performance. Using even a relatively simple numeric scale or above/average/below metric hides the student's actual trends and achievements/deficiencies without a thorough verbal description by previous instructors.
5	Objective standards and elimination of "average student" comparison except as designed within a program to establish trends. That is, submit scores and numerical assessments against a standard that is input to a database that will stratify a student within a group without instructor access to the averages.
6	Ease of use.
8	A sterile simulated event in which all injects could be controlled and evaluated objectively.
9	?
10	That tool already exists: either Access or SharePoint can provide the necessary capabilities. Basic database functions that allow data be to analyzed by any metric for which meta data exists are that main thing. Please, for the love of God, do not pay some crappy contractor to build an expensive, bloated, mostly useless system that will sit in a corner and get ignored.
11	Some type of system similar to FITREP grading average based on the Instructor's average. The problem with the current system is that there is an assumption that all IP's grade the same. As much as we attempt to standardize our grading procedures, there will always be some differences in grading criteria. If we could eliminate subjectivity by comparing PUI's performance against the IP's previous PUIs' performance and create an IP average for each T&R event, we would be able to compare apples to apples.

Response ID	Response Text
12	The hardest thing about ATFs and assigning numbered grades to specific categories is that it is all highly subjective. One Instructor's 3 is another Instructors 2. What constitutes those grades? Also with only 4 numbers there is no middle road. Is 2 slightly below average and 3 slightly above average? Again, it's subjective ... what is average? Perhaps there needs to be something that averages an instructor’s grades over time. Then you could look at an ATF and see "Oh look, this guy got a 3 in Situational Awareness, but his instructor's average is 2, so he must have done very well in that category." Something like you might see for the Marine Corps FitRep system. Right now with the numbers being subjective, I typically look to see if there's any glaring irregularities ... an ATF with straight 2s or straight 3s are pretty much the same to me. If I see 1s or 4s, I pay attention a little bit more. The Comments and Additional Remarks section is where an instructor must build a picture of the flight. It's not a place to continue instructing but to inform other instructors who weren't on the flight about how the flight went.
13	Comparative to a T/M/S population.
14	I am not sure what additional "tools" are out there that have not been debated already. "Assessing the performance of a PUI" (and the subsequent ATF) will always be subjective in nature. The reality is that only so many things can be numerically evaluated. I believe your assessment as the IP is based on (1) what you remember from when you flew the event as a PUI and (2) the other times you have instructed that event to other PUIs and how the PUIs stack-up. For a scored-shoot, one could include the score sheet, but I think all squadrons do that anyway. We conduct video debriefs after events, but I am not sure that linking an electronic ATF to a 1min "highlight" clip from the flight would be useful, maybe it would. For a while we messed around with putting the grade sheet on the ShareDrive so you could read how a peer group was performing, but not every 2301X or 2600X was given by the same IP to all PUIs (different scenarios, different grades). I think the current ATF works, provided you have a proactive IP corps that engages at the monthly IP board, and effective mentors that can then relay IP board results to the particular PUI.

Response ID	Response Text
18	We have this exact tool. It is the ATF's.
20	It should be able to pictorially depict trends in a PUIs training progression. I want a snap shot of a PUIs strengths and weaknesses that help me to focus my instruction where the specific PUI needs work.
21	An automatically updating chart that shows a peer group's performance compared to each other and compared to historical average for each ATF.
22	In my opinion there is no one additional tool that will aid the instructor to the point of making the current system any more effective than it is. When it comes down to it, the factors that determine whether or not a PUI will make it through the FRS and to the fleet lie above the level of the company grade IP. This is why you continue to see substandard individuals making through the course despite the objections of the instructor cadre. While the current grading system remains more ambiguous than the previously relied upon below/average/above scale, it is adequate enough to at least let the IP relate the performance of the individual to the standards established by the T&R manual.
23	Nothing of note at this time.
24	A way of tracking and comparing numerical average for PUIs across the fleet. A very brief synopsis of the PUIs performance for Squadron leadership that initials EVERY ATF.
25	It would be useful to have easy access to the mean performance on each flight. That would give an IP more insight as to where this student falls within the historical data.
26	I would prefer to go back to the above average, below average or average grading scale. This more clearly defined how a PUI was progressing. A lack of understanding of the grading scale now makes it difficult for students to receive a fair numerical grade from instructors. With this in mind, if a better understanding of the grading process was achieved, an excel spreadsheet, with associated graphs, that track students improvement and performance against other students would be helpful.
28	The concept of "Inter-Rater Reliability", IRR, is a novel one and well-intentioned. Unfortunately, it is scarcely applicable to ATF standardization. By that, I mean to say that we lose sight of the fact that assessing PUIs is an inherently subjective endeavor; to be so rigidly confined to making it objective is to betray the very nature of assessment. Like most things in the military, it stems from an attempt for uniformity and standardization, yet it is nigh impossible to holistically encapsulate a PUI's performance in a three-digit number.

Response ID	Response Text
	To that end, the remarks and additional comments sections of an ATF are where the real evaluation and assessment must take place. These subjective comments reinforce the subjective nature of the sortie. It is important to note that I am not advocating the abolishment of an ATF overall "score". In fact, it can be useful to compare PUIs with each other for a given sortie or stage, but only within that particular IP's metric. This sort of relativistic assessment is already seen in FITREP relative values. To then take an ATF score and compare it against some sort of mythical uniform standard is myopic and misleading.
29	-
30	Ability to compare one student against their peers across multiple squadrons or individual squadrons (subjective opinions would naturally be embedded). Ability to see an individual student's strengths & weaknesses across all stages/flights in one place. See remarks on #19.
31	Electronic ATFs that provide a real-time item average and would provide a list of comments for each performance item from the most recent 20 ATFs.
32	There should be some form of a baseline metric for each ATF. Our squadron has implemented a "threshold" and "expectation" for each ATF. They provide a tangible metric from which to base ATF grades on and allow an instructor to clearly state a message with the composite score at the bottom. The next step I would like to see taken would be a database that is updated (automatically linked via excel spreadsheet somehow) from which an IP could poll average grades for other instructors or students - similar to a RV on our FITREPs, or potentially poll certain events to see how the population as a whole performs.
38	ability to see trends for a specific mission or skill over time ability to see trends in weak or strong points over time
37	N/A
39	I would like to see a product that is capable of making the grading system more standardized or at least pull that information. As it stands the ATF is only helpful on the extreme sides of the scale. For example if a PUI receives multiple 4's my assumption based on the comments in regards to the lower portion of the ATF that he/she is doing extremely well, especially if the PUI is in the 2000 or 3000 level portion of the T&R. On the opposite side of the

Response ID	Response Text
	scale is either UNSAT or a 1. Both grades from my view point are negative and show a negative trend. A recap of thoughts; I believe it will be difficult to produce an accurate product based on the preference of the IP involved in the flight and his/her take on what the value of 1-4.
48	Compare avg. grades of PUIs across instructor's average. To take instructor bias out of grading. How stud compares to instructor avg.
41	Something that showed a trend of strengths and weaknesses of a PUI and also what sort of things the PUI has been exposed to. For example if during the CAS T&R events a PUI has never been exposed to or has shown weakness with 9-lines requiring multiple simultaneous HF, I could build a scenario to provide exposure and repetitions in the weak or new areas.
42	I'd like it to track weaknesses and identify to a crowd where shortfalls are popping up. Whether that's from airwork, to planning, to discussion items and studying. I'd like it in an easy presentable format. Similar to an NSS perhaps a system to show where a guy falls out compared to peers. Not to outcast him but to help catch them before they fall too far down a hole. An electronic system that can be accessed for all IP's to view would be much easier as well.
43	Standardization among pilots in stage fleet-wide
44	A tool that has ATF critical information directly reflected in the T&R.
45	Snapshot of all evaluated aspects from the current syllabus on one page, grades of all other PUIs in the unit in the same syllabus and record of those in the last year that could be shown after completing the ATF.
47	character
50	A trend indication would probably be the most important to me. Is a PUI or group of PUIs struggling in some areas while particularly strong in others?? I could use that information to tailor scenarios to help address deficiencies.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. WORD COUNT JAVA PROGRAM

```
package cs2173.swing;

import cs2173.collections.IgnoreCaseStringComparator;

import java.awt.BorderLayout;

import java.awt.Dimension;

import java.awt.Toolkit;

import java.awt.event.ActionEvent;

import java.awt.event.ActionListener;

import java.awt.event.WindowAdapter;

import java.awt.event.WindowEvent;

import java.awt.event.WindowListener;

import java.io.File;

import java.io.FileNotFoundException;

import java.io.FileWriter;

import java.io.IOException;

import java.util.Iterator;

import java.util.Scanner;

import java.util.SortedMap;

import java.util.TreeMap;

import javax.swing.JButton;

import javax.swing.JFileChooser;

import javax.swing.JFrame;

import javax.swing.JMenu;
```

```
import javax.swing.JMenuBar;
import javax.swing.JMenuItem;
import javax.swing.JOptionPane;
import javax.swing.JPanel;
import javax.swing.JScrollPane;
import javax.swing.JTextArea;
import javax.swing.SwingUtilities;
import javax.swing.UIManager;

/**
 * The User opens a text file and displays the counts of unique words in the
 * window. The user has the option of saving the counts to another text file.
 * <p>This version has an Exit button that prompts the user to confirm S/He
 * indeed wishes to exit the program. This happens when the user tries to close
 * the window directly via clicking 'X'.
 *
 * @version $Id: CountWordsGUI.java 170 2013-03-15 16:55:17Z ahbuss $
 * @author ahbuss
 */
public class CountWordsGUI extends JFrame implements Runnable {

 private JButton openButton;
 private JButton countButton;
 private JButton saveButton;
```

```
private JButton exitButton;
private JTextArea textArea;
private JFileChooser openFileChooser;
private JFileChooser saveFileChooser;
private SortedMap<String, Integer> wordCount;
private ExitActionListener exitActionListener;

public CountWordsGUI() {
super("Count Words");

// Changed to DO_NOTHINHG_ON_CLOSE to prevent window from
// closing without prompting user for confirmation.
// this.setDefaultCloseOperation(JFrame.EXIT_ON_CLOSE);
this.setDefaultCloseOperation(JFrame.DO_NOTHING_ON_CLOSE);
ExitWindowListener exitWindowListener = new ExitWindowListener();
this.addWindowListener(exitWindowListener);

// Instantiate the JTextArea where the counts will be displayed
// Wrap in a JScrollPane for scrolling
textArea = new JTextArea();
textArea.setEditable(false);
JScrollPane scrollPane = new JScrollPane(textArea);
this.getContentPane().add(scrollPane, BorderLayout.CENTER);
```

```
// The buttons will be in a single panel at the top of the window
 JPanel buttonPanel = new JPanel();

// Instantiate buttons
 openButton = new JButton("Open");
 countButton = new JButton("Count Words");
 saveButton = new JButton("Save Counts");
 exitButton = new JButton("Exit");

// Connect the ActionListeners to their respective buttons
 OpenActionListener openActionListener = new OpenActionListener();
 openButton.addActionListener(openActionListener);

 CountActionListener countActionListener = new CountActionListener();
 countButton.addActionListener(countActionListener);

 SaveActionListener saveActionListener = new SaveActionListener();
 saveButton.addActionListener(saveActionListener);

 exitActionListener = new ExitActionListener();
 exitButton.addActionListener(exitActionListener);

// Add each button to the buttonPanel and add the buttonPanel
// to the top (NORTH) of the ContentPane.
```

```
buttonPanel.add(openButton);
buttonPanel.add(countButton);
buttonPanel.add(saveButton);
buttonPanel.add(exitButton);

this.getContentPane().add(buttonPanel, BorderLayout.NORTH);

JMenuBar menuBar = new JMenuBar();

JMenu fileMenu = new JMenu("File");
menuBar.add(fileMenu);

JMenuItem openMenuItem = new JMenuItem("Open");
openMenuItem.addActionListener(openActionListener);
fileMenu.add(openMenuItem);

JMenuItem saveMenuItem = new JMenuItem("Save");
saveMenuItem.addActionListener(saveActionListener);
fileMenu.add(saveMenuItem);

fileMenu.addSeparator();
JMenuItem exitMenuItem = new JMenuItem("Exit");
exitMenuItem.addActionListener(exitActionListener);
fileMenu.add(exitMenuItem);
```

```
JMenu editMenu = new JMenu("Edit");
JMenuItem countWordsMenuItem = new JMenuItem("Count Words");
countWordsMenuItem.addActionListener(countActionListener);
editMenu.add(countWordsMenuItem);

JMenu helpMenu = new JMenu("Help");
AboutActionListener aboutActionListener = new AboutActionListener();
JMenuItem aboutMenuItem = new JMenuItem("About");
aboutMenuItem.addActionListener(aboutActionListener);
helpMenu.add(aboutMenuItem);

menuBar.add(editMenu);
menuBar.add(helpMenu);

this.setJMenuBar(menuBar);

// Instantiate the Ma that will contain the word counts.
this.wordCount = new TreeMap<String, Integer>(new IgnoreCaseStringComparator());
}

/**
* Sets the look-and-feel to the operating system being run using
```

```
* UIManager.setLookAndFeel().
*
* @throws a variety of exceptions from UIManager.setLookAndFeel() call
* @param args the command line arguments
*/
public static void main(String[] args) throws Throwable {
UIManager.setLookAndFeel(UIManager.getSystemLookAndFeelClassName());
CountWordsGUI countWordsGUI = new CountWordsGUI();
SwingUtilities.invokeLater(countWordsGUI);
}

/**
* Set size, location, and display. Centers the frame on the screen using
* Toolkit.getScreenSize().
*/
@Override
public void run() {
this.setSize(600, 500);
Toolkit toolkit = Toolkit.getDefaultToolkit();
Dimension screenSize = toolkit.getScreenSize();
int xLoc = (screenSize.width - this.getWidth()) / 2;
int yLoc = (screenSize.height - this.getHeight()) / 2;
this.setLocation(xLoc, yLoc);
this.setVisible(true);
```

```
}

private class OpenActionListener implements ActionListener {

/**
 * Open a JFileChooser for the user to select an input file. If a file
 * is selected, scan through the text and count the words. TODO: Move
 * the counting code to the CountActionListener
 *
 * @param e
 */
@Override
public void actionPerformed(ActionEvent e) {
 if (openFileChooser == null) {
 openFileChooser = new JFileChooser(System.getProperty("user.dir"));
 }
 int result = openFileChooser.showOpenDialog(CountWordsGUI.this);
 if (result == JFileChooser.APPROVE_OPTION) {
 File inputFile = openFileChooser.getSelectedFile();
 CountWordsGUI.this.setTitle(inputFile.getName() + " - Count Words");

 wordCount.clear();
 try {
  Scanner scanner = new Scanner(inputFile);
```

```
    while (scanner.hasNext()) {
    String line = scanner.nextLine();
    String[] splits = line.split("[\\s\\W\\d]+");
    for (String s : splits) {
//    This ignores empty words that somehow make it through
     if (!s.equals("")) {
     if (wordCount.containsKey(s)) {
      wordCount.put(s, wordCount.get(s) + 1);
     } else {
      wordCount.put(s, 1);
     }
     }
    }
    }
//   Added to clear textArea after opening another file
    textArea.setText("");
   } catch (FileNotFoundException ex) {
    throw new RuntimeException(ex);
   }
   }


  }
  }
```

```
public class CountActionListener implements ActionListener {

 /**
  * Display the wordCount contents in the JTextArea
  *
  * @param e
  */
 @Override
 public void actionPerformed(ActionEvent e) {
//  for (String key : wordCount.keySet()) {
  for (Iterator<String> iter = wordCount.keySet().iterator();
   iter.hasNext();) {
  String key = iter.next();
  textArea.append(key);
  textArea.append(" = ");
  textArea.append(wordCount.get(key).toString());
//  This is to eliminate the last empty line at the bottom
  if (iter.hasNext()) {
   textArea.append(System.getProperty("line.separator"));
  }
  }
  textArea.setCaretPosition(0);
 }
 }
```

```
private class SaveActionListener implements ActionListener {

/**
 * Prompt the user to enter a file to save the counts. Write the
 * contents of the JTextArea to the file and close.
 *
 * @param e
 */
@Override
public void actionPerformed(ActionEvent e) {
 if (saveFileChooser == null) {
 saveFileChooser = new JFileChooser();
 }
 int result = saveFileChooser.showSaveDialog(CountWordsGUI.this);
 if (result == JFileChooser.APPROVE_OPTION) {
 File outputFile = saveFileChooser.getSelectedFile();
 try {
  FileWriter outputFileWriter = new FileWriter(outputFile);
  outputFileWriter.write(textArea.getText());
  outputFileWriter.close();
 } catch (IOException ex) {
  throw new RuntimeException(ex);
 }
```

```
 }
}
}

private class ExitActionListener implements ActionListener {

/**
 * Prompt the user to confirm that they wish to exit. TODO: check that
 * there is an unsaved count. TODO: connect this ActionListener to when
 * the user clicks the close window icon.
 *
 * @param e
 */
@Override
public void actionPerformed(ActionEvent e) {
 int result = JOptionPane.showConfirmDialog(CountWordsGUI.this,
  "Are you really really sure?",
  "Are You Sure?", JOptionPane.OK_CANCEL_OPTION);
 if (result == JOptionPane.OK_OPTION) {
 System.exit(0);
 }
//  else if (result == JOptionPane.CANCEL_OPTION) {
//    JOptionPane.showMessageDialog(CountWordsGUI.this, "Exit Canceled by User");
```

```
//  }
 }
 }

 private class ExitWindowListener extends WindowAdapter implements WindowListener {

 /**
  * Calls exitActionListener.actionPerformed() to ensure that the same
  * behavior there is done when the window is closed.
  *
  * @param e
  */
 @Override
 public void windowClosing(WindowEvent e) {
  exitActionListener.actionPerformed(null);
 }
 }

 private class AboutActionListener implements ActionListener {

 @Override
 public void actionPerformed(ActionEvent e) {
  JOptionPane.showMessageDialog(rootPane, "Count Words GUI"
```

```
  + System.getProperty("line.separator")
  + "Counts unique words in a text file");
 }
 }
}
```

APPENDIX F. WORD COUNT RESULTS FROM FREE TEXT RESPONSE SURVEY QUESTIONS

a = 64

ability = 5

able = 2

abolishment = 1

about = 2

above = 5

access = 3

accessed = 1

accurate = 1

achieved = 1

achievements= 1

across = 4

actual = 1

Additional = 4

address = 1

adequate = 1

advocating = 1

affective = 1

after = 2

Again = 1

against = 5

aid = 1

airwork = 1

all = 9

allow = 2

already = 3

Also = 2

always = 2

am = 3

ambiguous = 1

among = 1

an = 24

analyzed = 1

and = 37

another = 1

any = 2

anyway = 1

apples = 2

applicable = 1

are = 6

areas = 2

around = 1

as = 9

aspects = 1

assessing = 3

assessment = 4

assessments = 1

assigning = 1

associated = 1

assumption = 2

at = 6

ATF = 19

ATFs = 3

attempt = 2

attention = 1

automatically= 2

average = 18

averages = 2

avg = 2

Awareness = 1

back = 1

base = 1

based = 4

baseline = 1

Basic = 1

be = 20

been = 3

before = 1

being = 1

believe = 2

below = 4

betray = 1

better = 1

bias = 1

bit = 1

bloated = 1

board = 2

Both = 1

bottom = 1

brief = 1

build = 3

but = 7

by = 6

cadre = 1

can = 5

capabilities = 1

capability = 1

capable = 1

CAS = 1

catch = 1

categories = 1

category = 1

certain = 1

character = 1

chart = 1

clearly = 2

clip = 1

cognitive = 1

comes = 1

comments = 6

company = 1

Comparative = 1

compare = 5

compared = 3

compares = 1

comparing = 2

comparison = 1

completing = 1

composite = 1

concept = 1

conduct = 1

confined = 1

constitutes = 1

continue = 2

contractor = 1

controlled = 1

corner = 1

Corps = 2

could = 9

course = 1

crappy = 1

create = 1

criteria = 1

critical = 1

crowd = 1

current = 5

d = 2

data = 3

database = 3

debated = 1

debriefs = 1

deficiencies = 2

defined = 1

depict = 1

description = 1

designed = 1
despite = 1
determine = 1
differences = 1
different = 2
difficult = 2
digit = 1
directly = 1
discussion = 1
do = 2
doing = 1
done = 1
down = 2
during = 1
each = 7
Ease = 1
easier = 1
easy = 2
eeach = 1
effective = 2
either = 2
electronic = 3
eliminate = 1
elimination = 1
embedded = 1
encapsulate = 1
end = 1
endeavor = 1
engages = 1
enough = 1
especially = 1
establish = 1
established = 1
evaluated = 3
evaluation = 1
even = 1
event = 4
events = 3
every = 2
exact = 1
example = 2
excel = 2
except = 1
exists = 2
expectation = 1
expensive = 1
exposed = 2
exposure = 1
extreme = 1
extrememly = 1
fact = 2
factors = 1
fair = 1
fall = 1
falls = 2
far = 1
FITREP = 3
FITREPs = 1
fleet = 3
flew = 1
flight = 6
flights = 1
focus = 1
for = 20
form = 1
format = 1
from = 10
FRS = 1
functions = 1
get = 1
give = 1
given = 2

glaring = 1

go = 1

God = 1

got = 1

grade = 4

grades = 9

grading = 9

graphs = 1

group = 4

guy = 2

hardest = 1

has = 5

have = 6

he = 2

help = 3

helpful = 2

her = 1

HF = 1

hides = 1

highlight = 1

highly = 1

him = 1

his = 2

historical = 2

hole = 1

holistically = 1

how = 6

I = 19

identify = 1

If = 7

ignored = 1

implemented = 1

important = 3

impossible = 1

improvement = 1

in = 28

include = 1

indication = 1

individual = 3

individuals = 1

inform = 1

information = 3

inherently = 1

initials = 1

injects = 1

input = 2

insight = 1

instructed = 1

instructing = 1

instruction = 1

instructor = 10

instructors = 6

intentioned = 1

Inter = 1

intructor = 1

involved = 1

IP = 15

IRR = 1

irregularities = 1

is = 35

it = 24

item = 2

items = 1

lack = 1

last = 1

leadership = 1

least = 2

let = 1

level = 2

lie = 1

like = 6

lines = 1

linked = 1

linking = 1

list = 1

little = 1

look = 3

lose = 1

love = 1

lower = 1

M = 1

main = 1

make = 1

makes = 1

making = 4

manual = 1

many = 1

Marine = 1

maybe = 1

me = 3

mean = 2

meaningful = 1

mentors = 1

message = 1

messed = 1

meta = 1

metric = 5

middle = 1

might = 1

military = 1

min = 1

mind = 1

misleading = 1

mission = 1

monthly = 1

more = 6

most = 4

mostly = 1

much = 3

multiple = 3

must = 3

my = 4

myopic = 1

mythical = 1

N = 1

naturally = 1

nature = 3

necessary = 1

needs = 2

negative = 2

never = 1

new = 1

next = 1

nigh = 1

no = 2

not = 9

note = 2

Nothing = 1

novel = 1

now = 2

NSS = 1

number = 1

numbered = 1

numbers = 2

numeric = 1

numerical = 3

numerically = 1

objections = 1

objective = 2

objectively = 1

of = 45

Oh = 1

on = 14

One = 7

only = 4

opinion = 1

opinions = 1

opposite = 1

or = 19

other = 8

others = 1

our = 3

out = 3

outcast = 1

over = 3

overall = 1

page = 1

particular = 2

particularly = 1

pay = 2

peer = 2

peers = 2

perfomance = 1

performance= 11

performing = 1

performs = 1

Perhaps = 2

pictorally = 1

picture = 1

pilots = 1

place = 3

planning = 1

Please = 1

point = 2

points = 1

poll = 2

popping = 1

population = 2

portion = 2

potentially = 1

prefer = 1

preference = 1

presentable = 1

pretty = 1

previous = 2

previously = 1

proactive = 1

probably = 1

problem = 1

proceedures = 1

process = 1

produce = 1

product = 2

program = 1

progressing = 1

progression = 1

provide = 5

provided = 1

psychomotor = 1

PUI = 14

PUIs = 14

pull = 1

putting = 1

R = 5

Rater = 1

read = 1

real = 2

reality = 1

recap = 1

receives = 1

recent = 1

recieve = 1

record = 1

reflected = 1

regarding = 1

regards = 1

reinforce = 1

relate = 1

relative = 1

relatively = 1

relativistic = 1

relay = 1

Reliability = 1

relied = 1

remains = 1

Remarks = 3

remember = 1

repetitions = 1

requiring = 1

results = 1

Right = 1

rigidily = 1

road = 1

RV = 1

s = 25

same = 4

say = 1

scale = 6

scarcely = 1

scenario = 1

scenarios = 2

score = 4

scored = 1

scores = 1

section = 1

sections = 1

see = 12

seen = 1

ShareDrive = 1

SharePoint = 1

she = 1

sheet = 2

shoot = 1

shortfalls = 1

shot = 1

should = 2

show = 2

showed = 1

shown = 2

shows = 1

side = 1

sides = 1

sight = 1

similar = 3

simple = 1

simultantous = 1

simulted = 1

sit = 1

Situational = 1

skill = 1

slightly = 2

snap = 1

Snapshot = 1

so = 4

some = 6

somehow = 1

something = 3

sort = 3

sortie = 2

specific = 3

spreadsheet = 2

Squadron = 2

squadrons = 3

stack = 1

stage = 2

stages = 1

standard = 3

standardization = 3

standardize = 1

standardized = 1

standards = 2

stands = 1

state = 1

stems = 1

step = 1

sterile = 1

straight = 2

stratify = 1

strenghts = 1

strengths = 2

strong = 2

struggling = 1

stud = 1

student = 7

students = 4

studying = 1

sub = 1

subjective = 8

subjectivity = 1

submit = 1

subsequent = 1

sure = 2

syllabus = 2

synopsis = 1

system = 9

T = 7

tailor = 1

take = 4

taken = 1

tangible = 1

than = 2

That = 42

The = 105

their = 1

them = 1

Then = 3

there = 8

These = 1

They = 2

thing = 2

things = 3

think = 2

this = 8

thorough = 1

those = 2

thoughts = 1

three = 1

threshold = 1

through = 2

time = 5

times = 1

to = 77

too = 1

tool = 4

tools = 1

track = 2

tracking = 1

training = 1

trend = 3

trends = 5

type = 1

typically = 1

understanding = 2

Unfortunately =1

uniform = 1

uniformity = 1

unit = 1

UNSAT = 1

up = 2

updated = 1

updating = 1

upon = 1

use = 2

useful = 3

useless = 1

Using = 1

value = 1

values = 1

verbal = 1

very = 3

via = 1

video = 1

view = 2

want = 1

was = 4

way = 1

we = 7

weak = 2

weakness = 1

weaknesses = 4

well = 4

went = 1

weren = 1

What = 6

when = 2

where = 6

whether = 2

which = 5

while = 3

who = 1

whole = 1

why = 1

wide = 1

will = 7

with = 10

within = 4

without = 2

work = 1

works = 1

would = 14

X = 2

year = 1

yet = 1

you = 8

your = 1

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Averweg, U. R. (2008). Decision support systems and decision-making processes. In F. Adam & P. Humphreys (Eds.), *Encyclopedia of decision making and decision support technologies* (pp. 218–224). Hershey, PA: IGI Global. doi:10.4018/978-1-59904-843-7.ch025

Beuschel, W. (2008). Dashboards for management. In F. Adam & P. Humphreys (Eds.), *Encyclopedia of decision making and decision support technologies* (pp. 116–123). Hershey, PA: IGI Global. doi:10.4018/978-1-59904-843-7.ch014

Beynon, M. J. (2005). A novel technique of object ranking and classification under ignorance: An application to the corporate failure risk problem. *European Journal of Operational Research*, *167*(2), 493–517. doi:10.1016/j.ejor.2004.03.016

Beynon, M. J. (2008a). Classification and ranking belief simplex. In F. Adam & P. Humphreys (Eds.), *Encyclopedia of decision making and decision support technologies* (pp. 76–83). Hershey, PA: IGI Global. doi:10.4018/978-1-59904-843-7.ch009

Beynon, M. J. (2008b). Promethee. In F. Adam & J.-C. Pomerol (Eds.), *Encyclopedia of decision making and decision support technologies* (pp. 743–750). Hershey, PA: IGI Global. doi:10.4018/978-1-59904-843-7.ch083

Beynon, M. J. (2008c). Qualitative comparative analysis. In F. Adam & J.-C. Pomerol (Eds.), *Encyclopedia of decision making and decision support technologies* (pp. 751–756). Hershey, PA: IGI Global. doi:10.4018/978-1-59904-843-7.ch084

Bloom, B. S. (1956). *Taxonomy of educational objectives*. New York: Longman.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. New York: Springer. Retrieved from http://link.springer.com.libproxy.nps.edu/book/10.1007/978-0-387-09506-6/page/1

Budrejko, T. A. (2009). *Instructor standardization: The key to excellence in Marine aviation* (Maser's thesis, United States Marine Corps Command and Staff College, Marine Corps University).

Commanding General Regional Contracting Office National Capital Region. (2014). *Contract solicitation number M00264-14-R-0022* (Vol. 1449). Quantico. Retrieved from https://www.fbo.gov/index?s=opportunity&mode=form&id=dc65047d410100ad1dcdd9dff36ced5f&tab=core&_cview=1

Dreyfus, S. E., & Dreyfus, H. L. (1980). A five-stage model of the mental activities in directed skill acquisition. Berkeley: University of California, Berkeley. Retrieved from http://handle.dtic.mil/100.2/ADA452068

Fenwick, M. (2010). Aviation Traning System: Winds of change in Marine Corps aviation training. *Marine Corps Gazette*, *94*(5), 52–58. Retrieved from http://pqasb.pqarchiver.com/mca-members/doc/221524187.html?FMT=PAGE&FMTS=ABS:FT:TG:PAGE&type=current&date=May+2010&author=Fenwick%2C+Mark+%22Skeeter%22&pub=Marine+Corps+Gazette&edition=&startpage=52-58&desc=Aviation+Training+System

Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. In *Proceedings Of The 2004 Conference on Designing Interactive Systems Processes, Practices, Methods, and Techniques - DIS '04* (pp. 261–268). New York: ACM Press. doi:10.1145/1013115.1013152

Gagne, R. M., & Briggs, L. J. (1979). *Principles of instructional design* (2nd ed.). New York: Holt, Rinehart and Winston.

Griffin, G. R. (1998). Predicting naval aviator flight training performance using multiple regression and an artificial neural network. *The International Journal of Aviation Psychology*, *8*(2), 121–135. doi:10.1207/s15327108ijap0802

Harlen, W., & James, M. (1997). Assessment and learning : Differences and relationships between formative and summative assessment. *Assessment in Education*, *4*(3), 365–378.

Hassenzahl, M., & Tractinsky, N. (2006). User experience: A research agenda. *Behaviour & Information Technology*, *25*(2), 91–97. doi:10.1080/01449290500330331

Headquarters United States Marine Corps. (2011a). AH-1W training and readiness manual. Washington, DC: Author

Headquarters United States Marine Corps. (2011b). *NAVMC 3500.14C Aviation training and readiness (T&R) program manual* (p. 242). Washington, DC: Author.

Hunter, D. R., & Burke, E. F. (2009). Predicting aircraft pilot training success: A meta-analysis of published research. *The International Journal of Aviation Psychology*, *4*(4), 297–313. doi:10.1207/s15327108ijap0404

Marine Light Attack Training Squadron 303 Operations Department. (2013). *Statistical process control brief (Microsoft PowerPoint)*. Camp Pendleton: Author.

Mersky, P. B. (1983). *U.S. Marine Corps aviation*. Baltimore: Nautical & Aviation Publishing Company.

Musick, D. W. (2006). A conceptual model for program evaluation in graduate medical education. *Academic Medicine : Journal of the Association of American Medical Colleges*, *81*(8), 759–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16868435

Naval Air Systems Command. (2011). *Navy tactics techniques and procedures* (3-22.5-AH1). Naval Air Systems Command.

Naval Air Training Command. (2007). *Chief of Naval Air Training instruction 1500.4G student naval air training and administration manual*. Corpus Christi: Chief of Naval Air Training.

Naval Aviation Schools Command. (2013). *Crew resource management instructor course student guide*. Pensacola: Naval Aviation Schools Command.

Office of the Deputy Commandant for Combat Development Integration. (2012). *2012 U.S. Marine Corps science and technology strategic plan*. Quantico. Retrieved from http://www.hqmc.marines.mil/Portals/160/Docs/USMC S_T Strat_Plan_2012_Final_31_ Jan.pdf

Paradice, D., & Davis, R. A. (2008). DSS and multiple perspectives of complex problems philosophical bases for perspective. In F. Adam & P. Humphreys (Eds.), *Encyclopedia of decision making and decision support technologies* (pp. 286–295). Hershey, PA. doi:10.4018/978-1-59904-843-7.ch033

Phillips, J. K., Shafer, J., Ross, K. G., & Cox, D. A. (2006). Behaviorally anchored rating scales for the assessment of tactical thinking mental models. Fort Knox: U.S. Army Research Institute. Retrieved from http://handle.dtic.mil/100.2/ADA452068

Power, B. (2008). Real options reasoning as a tool for managerial decision making. In F. Adam & J.-C. Pomerol (Eds.), *Encyclopedia of decision making and decision support technologies* (pp. 766–775). Hershey: IGI Global. doi:10.4018/978-1-59904-843-7.ch086

Rasmussen, J. (1986). *Information processing in human computer interaction: An approach to cognitive engineering*. San Diego: Elsevier.

Rickus, G. M., & Berkshire, J. R. (1968). Development of an aviation combat criterion: Preliminary report (Report 1047). Pensacola: Naval Aerospace Medical Institute. Retrieved from www.dtic.mil/cgi-bin/GetTRDoc?AD=AD0675214

Rudolph, J. W., Simon, R., Raemer, D. B., & Eppich, W. J. (2008). Debriefing as formative assessment: closing performance gaps in medical education. *Academic Emergency Medicine : Official Journal of the Society for Academic Emergency Medicine*, *15*(11), 1010–6. doi:10.1111/j.1553-2712.2008.00248.x

Shannon, R. H., & Waag, W. L. (1972). *Toward the development of a criterion for fleet effectiveness in the F-4 fighter community*. Pensacola: Naval Aerospace Medical Research Laboratory.

Stanley Jr., M. D. (1973). *A method for developing a criterion for combat performance of naval aviators*. Pensacola: Naval Aerospace Medical Research Laboratory.

Sutcliffe, A. (2010). *Designing for user engagement: Aeshetic and attractive user interfaces*. San Rafael: Morgan & Claypool.

Swezey, R. W. (1981). *Individual performance assessment: An approach to criterion-referenced test development*. Reston, VA: Reston Publishing Company.

Taras, M. (2005). Assessment :Summative and formative— Some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478.

U.S. Marine Corps. (2004). *Systems approach to training*. Quantico, VA: United States Marine Corps Combat Development Command.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California